(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 July 2002 (11.07.2002)

PCT

(10) International Publication Number
WO 02/053719 A2

(51) International Patent Classification[7]: C12N 9/00, C07K 14/47, C12N 15/12, 15/52, A61K 38/17, 38/43, C12Q 1/68, 1/25, 1/34, G01N 33/68, C07K 16/18, 16/40, C12N 15/00, A01K 67/027

(21) International Application Number: PCT/US02/00178

(22) International Filing Date: 4 January 2002 (04.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/260,085	4 January 2001 (04.01.2001)	US
60/268,554	13 February 2001 (13.02.2001)	US
60/269,111	14 February 2001 (14.02.2001)	US
60/271,211	23 February 2001 (23.02.2001)	US

(71) Applicant (*for all designated States except US*): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose CA 95123 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TANG, Y. Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue #306, Sunnyvale, CA 94086 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **STEWART, Elizabeth, A.** [US/US]; 1767 Monticello Road, San Mateo, CA 94402 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **KHAN, Farrah, A.** [IN/US]; 9445 Harrison Street, Des Plaines, IL 60016 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA 94024 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: CYTOSKELETON-ASSOCIATED PROTEINS

(57) Abstract: The invention provides human cytoskeleton-associated proteins (CSAP) and polynucleotides which identify and encode CSAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CSAP.

WO 02/053719 A2

CYTOSKELETON-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cytoskeleton-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative disorders, viral infections, and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of cytoskeleton-associated proteins.

BACKGROUND OF THE INVENTION

The cytoskeleton is a cytoplasmic network of protein fibers that mediate cell shape, structure, and movement. The cytoskeleton supports the cell membrane and forms tracks along which organelles and other elements move in the cytosol. The cytoskeleton is a dynamic structure that allows cells to adopt various shapes and to carry out directed movements. Major cytoskeletal fibers include the microtubules, the microfilaments, and the intermediate filaments. Motor proteins, including myosin, dynein, and kinesin, drive movement of or along the fibers. The motor protein dynamin drives the formation of membrane vesicles. Accessory or associated proteins modify the structure or activity of the fibers while cytoskeletal membrane anchors connect the fibers to the cell membrane.

Microtubules and Associated Proteins

Tubulins

Microtubules, cytoskeletal fibers with a diameter of about 24 nm, have multiple roles in the cell. Bundles of microtubules form cilia and flagella, which are whip-like extensions of the cell membrane that are necessary for sweeping materials across an epithelium and for swimming of sperm, respectively. Marginal bands of microtubules in red blood cells and platelets are important for these cells' pliability. Organelles, membrane vesicles, and proteins are transported in the cell along tracks of microtubules. For example, microtubules run through nerve cell axons, allowing bi-directional transport of materials and membrane vesicles between the cell body and the nerve terminal. Failure to supply the nerve terminal with these vesicles blocks the transmission of neural signals. Microtubules are also critical to chromosomal movement during cell division. Both stable and short-lived populations of microtubules exist in the cell.

Microtubules are polymers of GTP-binding tubulin protein subunits. Each subunit is a heterodimer of α- and β- tubulin, multiple isoforms of which exist. The hydrolysis of GTP is linked to the addition of tubulin subunits at the end of a microtubule. The subunits interact head to tail to form

protofilaments; the protofilaments interact side to side to form a microtubule. A microtubule is polarized, one end ringed with α-tubulin and the other with β-tubulin, and the two ends differ in their rates of assembly. Generally, each microtubule is composed of 13 protofilaments although 11 or 15 protofilament-microtubules are sometimes found. Cilia and flagella contain doublet microtubules. Microtubules grow from specialized structures known as centrosomes or microtubule-organizing centers (MTOCs). MTOCs may contain one or two centrioles, which are pinwheel arrays of triplet microtubules. The basal body, the organizing center located at the base of a cilium or flagellum, contains one centriole. Gamma tubulin present in the MTOC is important for nucleating the polymerization of α- and β- tubulin heterodimers but does not polymerize into microtubules.

Microtubule-Associated Proteins

Microtubule-associated proteins (MAPs) have roles in the assembly and stabilization of microtubules. One major family of MAPs, assembly MAPs, can be identified in neurons as well as non-neuronal cells. Assembly MAPs are responsible for cross-linking microtubules in the cytosol. These MAPs are organized into two domains: a basic microtubule-binding domain and an acidic projection domain. The projection domain is the binding site for membranes, intermediate filaments, or other microtubules. Based on sequence analysis, assembly MAPs can be further grouped into two types: Type I and Type II. Type I MAPs, which include MAP1A and MAP1B, are large, filamentous molecules that co-purify with microtubules and are abundantly expressed in brain and testes. Type I MAPs contain several repeats of a positively-charged amino acid sequence motif that binds and neutralizes negatively charged tubulin, leading to stabilization of microtubules. MAP1A and MAP1B are each derived from a single precursor polypeptide that is subsequently proteolytically processed to generate one heavy chain and one light chain.

Another light chain, LC3, is a 16.4 kDa molecule that binds MAP1A, MAP1B, and microtubules. It is suggested that LC3 is synthesized from a source other than the MAP1A or MAP1B transcripts, and that the expression of LC3 may be important in regulating the microtubule binding activity of MAP1A and MAP1B during cell proliferation (Mann, S.S. et al. (1994) J. Biol. Chem. 269:11492-11497).

Type II MAPs, which include MAP2a, MAP2b, MAP2c, MAP4, and Tau, are characterized by three to four copies of an 18-residue sequence in the microtubule-binding domain. MAP2a, MAP2b, and MAP2c are found only in dendrites, MAP4 is found in non-neuronal cells, and Tau is found in axons and dendrites of nerve cells. Alternative splicing of the Tau mRNA leads to the existence of multiple forms of Tau protein. Tau phosphorylation is altered in neurodegenerative disorders such as Alzheimer's disease, Pick's disease, progressive supranuclear palsy, corticobasal

degeneration, and familial frontotemporal dementia and Parkinsonism linked to chromosome 17. The altered Tau phosphorylation leads to a collapse of the microtubule network and the formation of intraneuronal Tau aggregates (Spillantini, M.G. and M. Goedert (1998) Trends Neurosci. 21:428-433).

The protein pericentrin is found in the MTOC and has a role in microtubule assembly.

Microfilaments and Associated Proteins

Actins

Microfilaments, cytoskeletal filaments with a diameter of about 7-9 nm, are vital to cell locomotion, cell shape, cell adhesion, cell division, and muscle contraction. Assembly and disassembly of the microfilaments allow cells to change their morphology. Microfilaments are the polymerized form of actin, the most abundant intracellular protein in the eukaryotic cell. Human cells contain six isoforms of actin. The three α-actins are found in different kinds of muscle, nonmuscle β-actin and nonmuscle γ-actin are found in nonmuscle cells, and another γ-actin is found in intestinal smooth muscle cells. G-actin, the monomeric form of actin, polymerizes into polarized, helical F-actin filaments, accompanied by the hydrolysis of ATP to ADP. Actin filaments associate to form bundles and networks, providing a framework to support the plasma membrane and determine cell shape. These bundles and networks are connected to the cell membrane. In muscle cells, thin filaments containing actin slide past thick filaments containing the motor protein myosin during contraction. A family of actin-related proteins exist that are not part of the actin cytoskeleton, but rather associate with microtubules and dynein.

Actin-Associated Proteins

Actin-associated proteins have roles in cross-linking, severing, and stabilization of actin filaments and in sequestering actin monomers. Several of the actin-associated proteins have multiple functions. Bundles and networks of actin filaments are held together by actin cross-linking proteins. These proteins have two actin-binding sites, one for each filament. Short cross-linking proteins promote bundle formation while longer, more flexible cross-linking proteins promote network formation. Calmodulin-like calcium-binding domains in actin cross-linking proteins allow calcium regulation of cross-linking. Group I cross-linking proteins have unique actin-binding domains and include the 30 kD protein, EF-1a, fascin, and scruin. Group II cross-linking proteins have a 7,000-MW actin-binding domain and include villin and dematin. Group III cross-linking proteins have pairs of a 26,000-MW actin-binding domain and include fimbrin, spectrin, dystrophin, ABP 120, and filamin.

Severing proteins regulate the length of actin filaments by breaking them into short pieces or by blocking their ends. Severing proteins include gCAP39, severin (fragmin), gelsolin, and villin. Capping proteins can cap the ends of actin filaments, but cannot break filaments. Capping proteins

include CapZ and tropomodulin. The proteins thymosin and profilin sequester actin monomers in the cytosol, allowing a pool of unpolymerized actin to exist. The actin-associated proteins tropomyosin, troponin, and caldesmon regulate muscle contraction in response to calcium.

Microtubule and actin filament networks cooperate in processes such as vesicle and organelle transport, cleavage furrow placement, directed cell migration, spindle rotation, and nuclear migration. Microtubules and actin may coordinate to transport vesicles, organelles, and cell fate determinants, or transport may involve targeting and capture of microtubule ends at cortical actin sites. These cytoskeletal systems may be bridged by myosin-kinesin complexes, myosin-CLIP170 complexes, formin-homology (FH) proteins, dynein, the dynactin complex, Kar9p, coronin, ERM proteins, and kelch repeat-containing proteins (for a review, see Goode, B.L. et al. (2000) Curr. Opin. Cell Biol. 12:63-71). The kelch repeat is a motif originally observed in the kelch protein, which is involved in formation of cytoplasmic bridges called ring canals. A variety of mammalian and other kelch family proteins have been identified. The kelch repeat domain is believed to mediate interaction with actin (Robinson, D.N. and L. Cooley (1997) J. Cell Biol. 138:799-810).

ADF/cofilins are a family of conserved 15-18 kDa actin-binding proteins that play a role in cytokinesis, endocytosis, and in development of embryonic tissues, as well as in tissue regeneration and in pathologies such as ischemia, oxidative or osmotic stress. LIM kinase 1 downregulates ADF (Carlier, M.F. et al. (1999) J. Biol. Chem. 274:33827-33830).

Intermediate Filaments and Associated Proteins

Intermediate filaments (IFs) are cytoskeletal fibers with a diameter of about 10 nm, intermediate between that of microfilaments and microtubules. IFs serve structural roles in the cell, reinforcing cells and organizing cells into tissues. IFs are particularly abundant in epidermal cells and in neurons. IFs are extremely stable, and, in contrast to microfilaments and microtubules, do not function in cell motility.

Five types of IF proteins are known in mammals. Type I and Type II proteins are the acidic and basic keratins, respectively. Heterodimers of the acidic and basic keratins are the building blocks of keratin IFs. Keratins are abundant in soft epithelia such as skin and cornea, hard epithelia such as nails and hair, and in epithelia that line internal body cavities. Mutations in keratin genes lead to epithelial diseases including epidermolysis bullosa simplex, bullous congenital ichthyosiform erythroderma (epidermolytic hyperkeratosis), non-epidermolytic and epidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, pachyonychia congenita, and white sponge nevus. Some of these diseases result in severe skin blistering. (See, e.g., Wawersik, M. et al. (1997) J. Biol. Chem. 272:32557-32565; and Corden L.D. and W.H. McLean (1996) Exp. Dermatol. 5:297-307.)

Type III IF proteins include desmin, glial fibrillary acidic protein, vimentin, and peripherin. Desmin filaments in muscle cells link myofibrils into bundles and stabilize sarcomeres in contracting muscle. Glial fibrillary acidic protein filaments are found in the glial cells that surround neurons and astrocytes. Vimentin filaments are found in blood vessel endothelial cells, some epithelial cells, and mesenchymal cells such as fibroblasts, and are commonly associated with microtubules. Vimentin filaments may have roles in keeping the nucleus and other organelles in place in the cell. Type IV IFs include the neurofilaments and nestin. Neurofilaments, composed of three polypeptides NF-L, NF-M, and NF-H, are frequently associated with microtubules in axons. Neurofilaments are responsible for the radial growth and diameter of an axon, and ultimately for the speed of nerve impulse transmission. Changes in phosphorylation and metabolism of neurofilaments are observed in neurodegenerative diseases including amylotrophic lateral sclerosis, Parkinson's disease, and Alzheimer's disease (Julien, J.P. and W.E. Mushynski (1998) Prog. Nucleic Acid Res. Mol. Biol. 61:1-23). Type V IFs, the lamins, are found in the nucleus where they support the nuclear membrane.

IFs have a central α-helical rod region interrupted by short nonhelical linker segments. The rod region is bracketed, in most cases, by non-helical head and tail domains. The rod regions of intermediate filament proteins associate to form a coiled-coil dimer. A highly ordered assembly process leads from the dimers to the IFs. Neither ATP nor GTP is needed for IF assembly, unlike that of microfilaments and microtubules.

IF-associated proteins (IFAPs) mediate the interactions of IFs with one another and with other cell structures. IFAPs cross-link IFs into a bundle, into a network, or to the plasma membrane, and may cross-link IFs to the microfilament and microtubule cytoskeleton. Microtubules and IFs are particularly closely associated. IFAPs include BPAG1, plakoglobin, desmoplakin I, desmoplakin II, plectin, ankyrin, filaggrin, and lamin B receptor.

Cytoskeletal-Membrane Anchors

Cytoskeletal fibers are attached to the plasma membrane by specific proteins. These attachments are important for maintaining cell shape and for muscle contraction. In erythrocytes, the spectrin-actin cytoskeleton is attached to the cell membrane by three proteins, band 4.1, ankyrin, and adducin. Defects in this attachment result in abnormally shaped cells which are more rapidly degraded by the spleen, leading to anemia. In platelets, the spectrin-actin cytoskeleton is also linked to the membrane by ankyrin; a second actin network is anchored to the membrane by filamin. In muscle cells the protein dystrophin links actin filaments to the plasma membrane; mutations in the dystrophin gene lead to Duchenne muscular dystrophy.

<u>Focal adhesions</u>

Focal adhesions are specialized structures in the plasma membrane involved in the adhesion of a cell to a substrate, such as the extracellular matrix. Focal adhesions form the connection between an extracellular substrate and the cytoskeleton, and affect such functions as cell shape, cell motility and cell proliferation. Transmembrane integrin molecules form the basis of focal adhesions. Upon ligand binding, integrins cluster in the plane of the plasma membrane. Cytoskeletal linker proteins such as the actin binding proteins α-actinin, talin, tensin, vinculin, paxillin, and filamin are recruited to the clustering site. Key regulatory proteins, such as Rho and Ras family proteins, focal adhesion kinase, and Src family members are also recruited. These events lead to the reorganization of actin filaments and the formation of stress fibers. These intracellular rearrangements promote further integrin-ECM interactions and integrin clustering. Thus, integrins mediate aggregation of protein complexes on both the cytosolic and extracellular faces of the plasma membrane, leading to the assembly of the focal adhesion. Many signal transduction responses are mediated via various adhesion complex proteins, including Src, FAK, paxillin, and tensin. (For a review, see Yamada, K.M. and B. Geiger, (1997) Curr. Opin. Cell Biol. 9:76-85.)

IFs are also attached to membranes by cytoskeletal-membrane anchors. The nuclear lamina is attached to the inner surface of the nuclear membrane by the lamin B receptor. Vimentin IFs are attached to the plasma membrane by ankyrin and plectin. Desmosome and hemidesmosome membrane junctions hold together epithelial cells of organs and skin. These membrane junctions allow shear forces to be distributed across the entire epithelial cell layer, thus providing strength and rigidity to the epithelium. IFs in epithelial cells are attached to the desmosome by plakoglobin and desmoplakins. The proteins that link IFs to hemidesmosomes are not known. Desmin IFs surround the sarcomere in muscle and are linked to the plasma membrane by paranemin, synemin, and ankyrin.

Motor Proteins

<u>Myosin-related Motor Proteins</u>

Myosins are actin-activated ATPases, found in eukaryotic cells, that couple hydrolysis of ATP with motion. Myosin provides the motor function for muscle contraction and intracellular movements such as phagocytosis and rearrangement of cell contents during mitotic cell division (cytokinesis). The contractile unit of skeletal muscle, termed the sarcomere, consists of highly ordered arrays of thin actin-containing filaments and thick myosin-containing filaments. Crossbridges form between the thick and thin filaments, and the ATP-dependent movement of myosin heads within the thick filaments pulls the thin filaments, shortening the sarcomere and thus the muscle fiber.

Myosins are composed of one or two heavy chains and associated light chains. Myosin heavy chains contain an amino-terminal motor or head domain, a neck that is the site of light-chain binding,

and a carboxy-terminal tail domain. The tail domains may associate to form an α-helical coiled coil. Conventional myosins, such as those found in muscle tissue, are composed of two myosin heavy-chain subunits, each associated with two light-chain subunits that bind at the neck region and play a regulatory role. Unconventional myosins, believed to function in intracellular motion, may contain either one or two heavy chains and associated light chains. There is evidence for about 25 myosin heavy chain genes in vertebrates, more than half of them unconventional.

Dynein-related Motor Proteins

Dyneins are (-) end-directed motor proteins which act on microtubules. Two classes of dyneins, cytosolic and axonemal, have been identified. Cytosolic dyneins are responsible for translocation of materials along cytoplasmic microtubules, for example, transport from the nerve terminal to the cell body and transport of endocytic vesicles to lysosomes. As well, viruses often take advantage of cytoplasmic dyneins to be transported to the nucleus and establish a successful infection (Sodeik, B. et al. (1997) J. Cell Bio. 136:1007-1021). Virion proteins of herpes simplex virus 1, for example, interact with the cytoplasmic dynein intermediate chain (Ye, G.J. et al. (2000) J. Virol. 74:1355-1363). Cytoplasmic dyneins are also reported to play a role in mitosis. Axonemal dyneins are responsible for the beating of flagella and cilia. Dynein on one microtubule doublet walks along the adjacent microtubule doublet. This sliding force produces bending that causes the flagellum or cilium to beat. Dyneins have a native mass between 1000 and 2000 kDa and contain either two or three force-producing heads driven by the hydrolysis of ATP. The heads are linked via stalks to a basal domain which is composed of a highly variable number of accessory intermediate and light chains. Cytoplasmic dynein is the largest and most complex of the motor proteins.

Kinesin-related Motor Proteins

Kinesins are (+) end-directed motor proteins which act on microtubules. The prototypical kinesin molecule is involved in the transport of membrane-bound vesicles and organelles. This function is particularly important for axonal transport in neurons. Kinesin is also important in all cell types for the transport of vesicles from the Golgi complex to the endoplasmic reticulum. This role is critical for maintaining the identity and functionality of these secretory organelles.

Kinesins define a ubiquitous, conserved family of over 50 proteins that can be classified into at least 8 subfamilies based on primary amino acid sequence, domain structure, velocity of movement, and cellular function. (Reviewed in Moore, J.D. and S.A. Endow (1996) Bioessays 18:207-219; and Hoyt, A.M. (1994) Curr. Opin. Cell Biol. 6:63-68.) The prototypical kinesin molecule is a heterotetramer comprised of two heavy polypeptide chains (KHCs) and two light polypeptide chains (KLCs). The KHC subunits are typically referred to as "kinesin." KHC is about 1000 amino acids in

length, and KLC is about 550 amino acids in length. Two KHCs dimerize to form a rod-shaped molecule with three distinct regions of secondary structure. At one end of the molecule is a globular motor domain that functions in ATP hydrolysis and microtubule binding. Kinesin motor domains are highly conserved and share over 70% identity. Beyond the motor domain is an α-helical coiled-coil region which mediates dimerization. At the other end of the molecule is a fan-shaped tail that associates with molecular cargo. The tail is formed by the interaction of the KHC C-termini with the two KLCs.

Members of the more divergent subfamilies of kinesins are called kinesin-related proteins (KRPs), many of which function during mitosis in eukaryotes (Hoyt, supra). Some KRPs are required for assembly of the mitotic spindle. In vivo and in vitro analyses suggest that these KRPs exert force on microtubules that comprise the mitotic spindle, resulting in the separation of spindle poles. Phosphorylation of KRP is required for this activity. Failure to assemble the mitotic spindle results in abortive mitosis and chromosomal aneuploidy, the latter condition being characteristic of cancer cells. In addition, a unique KRP, centromere protein E, localizes to the kinetochore of human mitotic chromosomes and may play a role in their segregation to opposite spindle poles.

Dynamin-related Motor Proteins

Dynamin is a large GTPase motor protein that functions as a "molecular pinchase," generating a mechanochemical force used to sever membranes. This activity is important in forming clathrin-coated vesicles from coated pits in endocytosis and in the biogenesis of synaptic vesicles in neurons. Binding of dynamin to a membrane leads to dynamin's self-assembly into spirals that may act to constrict a flat membrane surface into a tubule. GTP hydrolysis induces a change in conformation of the dynamin polymer that pinches the membrane tubule, leading to severing of the membrane tubule and formation of a membrane vesicle. Release of GDP and inorganic phosphate leads to dynamin disassembly. Following disassembly the dynamin may either dissociate from the membrane or remain associated to the vesicle and be transported to another region of the cell. Three homologous dynamin genes have been discovered, in addition to several dynamin-related proteins. Conserved dynamin regions are the N-terminal GTP-binding domain, a central pleckstrin homology domain that binds membranes, a central coiled-coil region that may activate dynamin's GTPase activity, and a C-terminal proline-rich domain that contains several motifs that bind SH3 domains on other proteins. Some dynamin-related proteins do not contain the pleckstrin homology domain or the proline-rich domain. (See McNiven, M.A. (1998) Cell 94:151-154; Scaife, R.M. and R.L. Margolis (1997) Cell. Signal. 9:395-401.)

The cytoskeleton is reviewed in Lodish, H. et al. (1995) Molecular Cell Biology, Scientific

American Books, New York NY.

The discovery of new cytoskeleton-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative disorders, viral infections, and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of cytoskeleton-associated proteins.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, cytoskeleton-associated proteins, referred to collectively as "CSAP" and individually as "CSAP-1," "CSAP-2," "CSAP-3," "CSAP-4," "CSAP-5," "CSAP-6," "CSAP-7," "CSAP-8," "CSAP-9," "CSAP-10," "CSAP-11," "CSAP-12," "CSAP-13," "CSAP-14," "CSAP-15," "CSAP-16," "CSAP-17," and "CSAP-18." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-18. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:19-36.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least

90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CSAP, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an

agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CSAP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CSAP, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one

test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising

a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CSAP" refers to the amino acid sequences of substantially purified CSAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CSAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CSAP either by directly interacting with CSAP or by acting on components of the biological pathway in which CSAP participates.

An "allelic variant" is an alternative form of the gene encoding CSAP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CSAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CSAP or a polypeptide with at least one functional characteristic of CSAP. Included within this definition are

polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CSAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CSAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CSAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CSAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CSAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CSAP either by directly interacting with CSAP or by acting on components of the biological pathway in which CSAP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CSAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and

keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once

introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic CSAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CSAP or fragments of CSAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an

exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of CSAP or the polynucleotide encoding CSAP which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:19-36 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:19-36, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:19-36 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:19-36 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:19-36 and the region of SEQ ID NO:19-36 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-18 is encoded by a fragment of SEQ ID NO:19-36. A fragment of SEQ ID NO:1-18 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-18. For example, a fragment of SEQ ID NO:1-18 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-18. The precise length of a fragment of SEQ ID NO:1-18 and the region of SEQ ID NO:1-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two

or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of

the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CSAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of CSAP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides,

polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of CSAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CSAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an CSAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CSAP.

"Probe" refers to nucleic acid sequences encoding CSAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous

nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially

complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CSAP, nucleic acids encoding CSAP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For

example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host

by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human cytoskeleton-associated proteins (CSAP), the polynucleotides encoding CSAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative disorders, viral infections, and neurological disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are cytoskeleton-associated proteins. For example, SEQ ID NO:5 is 94% identical to dog Band 4.1-like 5 protein (GenBank ID g8979743) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.6e-264, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a Band 4.1 family FERM domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:5 is a Band 4.1

family protein. In an alternative example, SEQ ID NO:7 is 95% identical to human beta-tubulin (GenBank ID g1805274) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.4 e-227, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:7 also contains a tubulin/Ftsz family domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:7 is a tubulin. In an alternative example, SEQ ID NO:11 is 80% identical, from residue M1 to residue G529, to Mus musculus type II cytokeratin (GenBank ID g6092075) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.5e-213, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:11 also contains an intermediate filament domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:11 is an intermediate filament protein. In an alternative example, SEQ ID NO:17 is 90% identical, from residue M1 to residue I888, to Mus musculus POSH protein (GenBank ID g3002588) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also contains an SH3 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:17 is an SH3-containing protein. SEQ ID NO:1-4, SEQ ID NO:6, SEQ ID NO:8-10, SEQ ID NO:12-16 and SEQ ID NO:18 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-18 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide

sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:19-36 or that distinguish between SEQ ID NO:19-36 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_$XXXXXX$_N_1_N_2_$YYYYY$_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses CSAP variants. A preferred CSAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CSAP amino acid sequence, and which contains at least one functional or structural characteristic of CSAP.

The invention also encompasses polynucleotides which encode CSAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36, which encodes CSAP. The polynucleotide sequences of SEQ ID NO:19-36, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CSAP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CSAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-

36 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:19-36. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CSAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding CSAP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding CSAP, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding CSAP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding CSAP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CSAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CSAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CSAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CSAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CSAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CSAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CSAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CSAP and CSAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CSAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:19-36 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CSAP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent

to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CSAP may be cloned in recombinant DNA molecules that direct expression of CSAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CSAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CSAP-encoding sequences for a variety of purposes including, but

not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CSAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CSAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CSAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CSAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.)

The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active CSAP, the nucleotide sequences encoding CSAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CSAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CSAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CSAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CSAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CSAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO

J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding CSAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CSAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CSAP into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CSAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CSAP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CSAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CSAP. Transcription of sequences encoding CSAP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al.

(1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CSAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CSAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CSAP in cell lines is preferred. For example, sequences encoding CSAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g.,

Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CSAP is inserted within a marker gene sequence, transformed cells containing sequences encoding CSAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CSAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CSAP and that express CSAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CSAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CSAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CSAP include

oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CSAP, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CSAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CSAP may be designed to contain signal sequences which direct secretion of CSAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CSAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CSAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CSAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion

proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CSAP encoding sequence and the heterologous protein sequence, so that CSAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CSAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

CSAP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CSAP. At least one and up to a plurality of test compounds may be screened for specific binding to CSAP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CSAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CSAP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CSAP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing CSAP or cell membrane fractions which contain CSAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CSAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CSAP, either in solution or affixed to a solid support, and detecting the binding of CSAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.

Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

CSAP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CSAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CSAP activity, wherein CSAP is combined with at least one test compound, and the activity of CSAP in the presence of a test compound is compared with the activity of CSAP in the absence of the test compound. A change in the activity of CSAP in the presence of the test compound is indicative of a compound that modulates the activity of CSAP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising CSAP under conditions suitable for CSAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CSAP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CSAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CSAP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding CSAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CSAP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CSAP, e.g., by secreting CSAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CSAP and cytoskeleton-associated proteins. In addition, examples of tissues expressing CSAP can be found in Table 6. Therefore, CSAP appears to play a role in cell proliferative disorders, viral infections, and neurological disorders. In the treatment of disorders associated with increased CSAP expression or activity, it is desirable to decrease the expression or activity of CSAP. In the treatment of disorders associated with decreased CSAP expression or activity, it is desirable to increase the expression or activity of CSAP.

Therefore, in one embodiment, CSAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CSAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a viral infection such as those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornoviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever), retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses

(encephalitis, rubella); and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, a prion disease including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder.

In another embodiment, a vector capable of expressing CSAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CSAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CSAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CSAP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CSAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CSAP including, but not limited to, those listed above.

In a further embodiment, an antagonist of CSAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CSAP. Examples of such disorders include, but are not limited to, those cell proliferative disorders, viral infections, and neurological disorders described above. In one aspect, an antibody which specifically binds CSAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CSAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CSAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CSAP including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CSAP may be produced using methods which are generally known in the art. In particular, purified CSAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CSAP. Antibodies to CSAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CSAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CSAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of CSAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CSAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma

technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CSAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CSAP may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CSAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CSAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CSAP. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CSAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a

determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CSAP epitopes, represents the average affinity, or avidity, of the antibodies for CSAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CSAP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CSAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CSAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CSAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CSAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CSAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CSAP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other

gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CSAP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in CSAP expression or regulation causes disease, the expression of CSAP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CSAP are treated by constructing mammalian expression vectors encoding CSAP and introducing these vectors by mechanical means into CSAP-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CSAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors

(Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CSAP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CSAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CSAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CSAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.

Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4^{+} T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CSAP to cells which have one or more genetic abnormalities with respect to the expression of CSAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CSAP to target cells which have one or more genetic abnormalities with respect to the expression of CSAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CSAP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of

ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CSAP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CSAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CSAP-coding RNAs and the synthesis of high levels of CSAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of CSAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CSAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CSAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CSAP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CSAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CSAP may be therapeutically useful, and in the treatment of disorders associated with decreased CSAP expression or activity, a compound which specifically promotes expression of the

polynucleotide encoding CSAP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CSAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CSAP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CSAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and

monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CSAP, antibodies to CSAP, and mimetics, agonists, antagonists, or inhibitors of CSAP.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CSAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CSAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for

administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CSAP or fragments thereof, antibodies of CSAP, and agonists, antagonists or inhibitors of CSAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CSAP may be used for the diagnosis of disorders characterized by expression of CSAP, or in assays to monitor patients being treated with CSAP or agonists, antagonists, or inhibitors of CSAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CSAP include methods which utilize the antibody and a label to detect CSAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification,

and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CSAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CSAP expression. Normal or standard values for CSAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CSAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CSAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CSAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CSAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CSAP, and to monitor regulation of CSAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CSAP or closely related molecules may be used to identify nucleic acid sequences which encode CSAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CSAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CSAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:19-36 or from genomic sequences including promoters, enhancers, and introns of the CSAP gene.

Means for producing specific hybridization probes for DNAs encoding CSAP include the cloning of polynucleotide sequences encoding CSAP or CSAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels,

such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CSAP may be used for the diagnosis of disorders associated with expression of CSAP. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a viral infection such as those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornoviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever), retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella); and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, a prion disease including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias,

paranoid psychoses, postherpetic neuralgia, and Tourette's disorder. The polynucleotide sequences encoding CSAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CSAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CSAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CSAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CSAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CSAP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CSAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ

preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CSAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CSAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CSAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CSAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CSAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of CSAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid

quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CSAP, fragments of CSAP, or antibodies specific for CSAP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental

compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl

sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CSAP to quantify the levels of CSAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid

residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in <u>DNA Microarrays: A Practical Approach</u>, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CSAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent <u>in situ</u> hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, <u>supra</u>, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CSAP on a

physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CSAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CSAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CSAP, or fragments thereof, and washed. Bound CSAP is then detected by methods well known in the art. Purified CSAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CSAP specifically compete with a test compound for binding CSAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CSAP.

In additional embodiments, the nucleotide sequences which encode CSAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such

properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/260,085, U.S. Ser. No. 60/268,554, U.S. Ser. No. 60/269,111, and U.S. Ser. No. 60/271,211 are expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs

were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold

parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:19-36. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative cytoskeleton-associated proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode cytoskeleton-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for cytoskeleton-associated proteins. Potential cytoskeleton-associated proteins were also identified by homology to Incyte cDNA sequences that had been annotated as cytoskeleton-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. **Assembly of Genomic Sequence Data with cDNA Sequence Data**

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant

stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CSAP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:19-36 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:19-36 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:24 was mapped to chromosome 18 within the interval from 40.4 to 42.7 centiMorgans. SEQ ID NO:31 was mapped to chromosome 1 within the interval from the p-terminus to 16.40 centiMorgans. SEQ ID NO:33 was mapped to chromosome 19 within the interval from 19.1 to 35.5 centiMorgans. SEQ ID NO:25 was mapped to chromosome 6 within the interval from the p-terminus to 14.2 centiMorgans, to chromosome 16 within the interval from 44.3 to 45.4 centiMorgans, to chromosome 6 within the interval from 42.0 to 44.9 centiMorgans, and to chromosome 2 within the interval from 120.8 to 134.1 centiMorgans. More than one map location is reported for SEQ ID NO:25, indicating that sequences having different map locations were assembled into a single cluster. This situation occurs, for example, when sequences having strong similarity, but not complete identity, are assembled into a single cluster.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a

gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score x Percent Identity}}{5 \text{ x minimum } \{\text{length(Seq. 1), length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding CSAP are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or

urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding CSAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CSAP Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the

concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:19-36 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a

SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

<u>Tissue or Cell Sample Preparation</u>

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by <u>in vitro</u> transcription from non-coding yeast genomic DNA. After incubation at 37° C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85° C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

<u>Microarray Preparation</u>

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60° C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore,

are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the CSAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CSAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CSAP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CSAP-encoding transcript.

XII. Expression of CSAP

Expression and purification of CSAP is achieved using bacterial or virus-based expression systems. For expression of CSAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CSAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CSAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly

known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CSAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CSAP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CSAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CSAP obtained by these methods can be used directly in the assays shown in Examples XVI and XVII where applicable.

XIII. Functional Assays

CSAP function is assessed by expressing the sequences encoding CSAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of

fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CSAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CSAP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CSAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of CSAP Specific Antibodies

CSAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CSAP amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CSAP activity by, for example, binding the peptide or CSAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring CSAP Using Specific Antibodies

Naturally occurring or recombinant CSAP is substantially purified by immunoaffinity chromatography using antibodies specific for CSAP. An immunoaffinity column is constructed by covalently coupling anti-CSAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CSAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CSAP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CSAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CSAP is collected.

XVI. Identification of Molecules Which Interact with CSAP

CSAP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CSAP, washed, and any wells with labeled CSAP complex are assayed. Data obtained using different concentrations of CSAP are used to calculate values for the number, affinity, and association of CSAP with the candidate molecules.

Alternatively, molecules interacting with CSAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

CSAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of CSAP Activity

A microtubule motility assay for CSAP measures motor protein activity. In this assay, recombinant CSAP is immobilized onto a glass slide or similar substrate. Taxol-stabilized bovine brain microtubules (commercially available) in a solution containing ATP and cytosolic extract are perfused onto the slide. Movement of microtubules as driven by CSAP motor activity can be visualized and quantified using video-enhanced light microscopy and image analysis techniques. CSAP activity is directly proportional to the frequency and velocity of microtubule movement.

Alternatively, an assay for CSAP measures the formation of protein filaments in vitro. A solution of CSAP at a concentration greater than the "critical concentration" for polymer assembly is applied to carbon-coated grids. Appropriate nucleation sites may be supplied in the solution. The grids are negative stained with 0.7% (w/v) aqueous uranyl acetate and examined by electron microscopy. The appearance of filaments of approximately 25 nm (microtubules), 8 nm (actin), or 10 nm (intermediate filaments) is a demonstration of protein activity.

In another alternative, CSAP activity is measured by the binding of CSAP to protein filaments. ^{35}S-Met labeled CSAP sample is incubated with the appropriate filament protein (actin, tubulin, or intermediate filament protein) and complexed protein is collected by immunoprecipitation using an antibody against the filament protein. The immunoprecipitate is then run out on SDS-PAGE and the amount of CSAP bound is measured by autoradiography.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
5566074	1	5566074CD1	19	5566074CB1
5679814	2	5679814CD1	20	5679814CB1
7472735	3	7472735CD1	21	7472735CB1
7131221	4	7131221CD1	22	7131221CB1
7480551	5	7480551CD1	23	7480551CB1
3315870	6	3315870CD1	24	3315870CB1
7484690	7	7484690CD1	25	7484690CB1
7612559	8	7612559CD1	26	7612559CB1
4940751	9	4940751CD1	27	4940751CB1
7946761	10	7946761CD1	28	7946761CB1
3288747	11	3288747CD1	29	3288747CB1
8200016	12	8200016CD1	30	8200016CB1
3291962	13	3291962CD1	31	3291962CB1
1234259	14	1234259CD1	32	1234259CB1
1440608	15	1440608CD1	33	1440608CB1
3413610	16	3413610CD1	34	3413610CB1
3276394	17	3276394CD1	35	3276394CB1
7602049	18	7602049CD1	36	7602049CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	5566074CD1	g2200	1.8e-196	[Sus scrofa] Tubulin-tyrosine ligase Ersfeld, K. et al. (1993) J. Cell Biol. 120:725-732
2	5679814CD1	g2645229	1.5e-37	[Plectonema boryanum] Kinesin light chain Celerin, M. et al. (1997) DNA Cell Biol. 16:787-795
3	7472735CD1	g710551	4.1e-31	[Mus musculus] Ankyrin 3 Peters, L.L. et al. (1995) J. Cell Biol. 130:313-330
4	7131221CD1	g9945010	2.2e-95	[Mus musculus] RING-finger protein MURF Spencer, J.A. et al. (2000) J. Cell Biol. 150:771-784
5	7480551CD1	g8979743	1.6e-264	[Canis familiaris] Band 4.1-like 5 protein
6	3315870CD1	g1167996	7.8e-50	[Homo sapiens] ankyrin G119 Devarajan,P. et al. (1996) J. Cell Biol. 133 (4), 819-830
7	7484690CD1	g1805274	5.4e-227	[Homo sapiens] beta-tubulin van Geel,M. et al. (2000) Cytogenet Cell Genet. 2000;88(3-4):316-21
8	7612559CD1	g64402	5.0e-9	[Torpedo californica] type III intermediate filament
9	4940751CD1	g1419370	4.3e-69	[Zea mays] actin depolymerizing factor Lopez,I., et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:7415-7420
10	7946761CD1	g1841966	7.0e-08	[Rattus norvegicus] ankyrin
11	3288747CD1	g6092075	2.5e-213	[Mus musculus] type II cytokeratin
12	8200016CD1	g6636340	0.0	[Rattus norvegicus] myosin heavy chain Myr 8
13	3291962CD1	g12248771	2.2e-278	[Homo sapiens] (AB014736) SMAP-1b smooth muscle cell associated protein
14	1234259CD1	g10312104	5.0e-224	[Mus musculus] SMAR1 matrix/scaffold-associated region binding protein
15	1440608CD1	g4050093	0.0	[Mus musculus] ankyrin-related NG28
16	3413610CD1	g2104558	7.4e-278	[Rattus norvegicus] CCA3 Hayashi, Y. et al. (1997) FEBS Lett. 406:147-150
17	3276394CD1	g3002588	0.0	[Mus musculus] POSH Tapon, N. et al. (1998) EMBO J. 17:1395-1404 (1998)
18	7602049CD1	g5441367	9.8e-143	[Homo sapiens] ZASP protein Faulkner, G. et al. (1999) J. Cell Biol. 146:465-476

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	5566074CD1	377	S76 S123 S303 T83 T223	N10 N276	Signal peptide: M1-R32	SPScan
					Tubulin-tyrosine ligase PD008766: M1-T304, K198-L377	BLAST-PRODOM
2	5679814CD1	696	S39 S120 S149 S209 S231 S291 S419 S674 S680 T195 T228 T341 T392 T461 T545 T629	N142 N304	TPR Domain: A585-S618, A501-A534, A543-S576, A351-A384, A459-V492, G627-E660, A309-A342, A393-A426	HMMER-PFAM
					Kinesin light chain repeat proteins BL01160: S485-E525	BLIMPS-BLOCKS
					Leucine zipper pattern: L528-L549	MOTIFS
3	7472735CD1	1050	S122 S134 S149 S170 S245 S299 S368 S396 S418 S445 S458 S477 S555 S612 S657 S658 S704 S902 S970 T137 T162 T353 T663 T692 T740 T878 T909 T916 T952 T980 T1023 T1041	N331 N490 N738 N837 N965	Vacuolar sorting protein 9 (VPS9) domain: I264-A369	HMMER-PFAM
					Ankyrin repeats: S809-N841, K842-K874, R462-Y494, K564-R596, N528-I560, D743-A775, D776-L808, H495-N527	HMMER-PFAM
					Transmembrane domains: G77-N102, V851-S868 N-terminus is non-cytosolic	TMAP
					ATP/GTP-binding site motif A (P-loop): A945-T952	MOTIFS
4	7131221CD1	326	S80 S112 S118 S174 S228 S238 S262 S301 T281	N2	Zinc finger, C3HC4 type (RING finger): C26-C50	HMMER-PFAM
					Zinc finger, C3HC4 type: C42-C50	BLIMPS-BLOCKS
					Zinc finger, C3HC4 type (RING finger), signature: K22-G91	ProfileScan
					Zinc finger, C3HC4 type (RING finger), signature: C42-A51	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7480551CD1	505	S13 S48 S103 S149 S333 S348 S372 S376 T44 T272 T355 T358 T365 T387 Y245	N397 N403 N475	FERM domain (Band 4.1 family): C45-H235	HMMER-PFAM
					Band 4.1 family domain signature 1: W97-E126	MOTIFS
					Band 4.1 family domain signature 2: W205-M234	MOTIFS
					Band 4.1 family domain proteins BL00660: G52-I104, R136-D175, Q215-E258, F266-D289, F301-F323	BLIMPS-BLOCKS
					Band 4.1 family domain signatures: K102-D146	ProfileScan
					Band 4.1 family domain signatures : G210-E258	ProfileScan
					ERM family signature PR00661: Q107-E126, G150-L171, K238-E258, Y347-E368, S56-H75	BLIMPS-PRINTS
					Band 4.1 protein family signature PR00935: L76-F88, L141-C154, C154-Y174, Q215-G231	BLIMPS-PRINTS
					Cytoskeleton structural protein, phosphatase, hydrolase, tyrosine phosphorylation, Band PD000961: C45-D233	BLAST-PRODOM
					Cytoskeleton structural protein, phosphatase, hydrolase, tyrosine phosphorylation, Band PD014063: M234-K388	BLAST-PRODOM
					Band 4: DM00609 | P29074 | 19-463: I43-G410 DM00609 | P11171 | 200-623: C45-P437 DM00609 | P52963 | 2-423: C45-H442 DM00609 | P11434 | 183-612: C45-S399	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	3315870CD1	367	S39 S84 S136 S361 T200 T365	N134	Ank repeat: T7-N72, L75-K367	HMMER_PFAM
					Transmembrane domain: R42-K70 K171-V199 N terminus non-cytosolic	TMAP
					Ank repeat proteins PF00023: L177-L192, G339-R348	BLIMPS_PFAM
					Ank repeat protein PD00078B: D336-R348	BLIMPS_PRODOM
					F34D10.6 PROTEIN PD020606: A176-N328	BLAST_PRODOM
7	7484690CD1	435	S76 S116 S173 S382 T179 T215 T222 T275 T286 T388 T409	N185 N338 N371	Tubulin/FtsZ family: M1-Q423	HMMER_PFAM
					Tubulin subunits alpha, beta, and gamma proteins BL00227: R2-G35, H51-G105, E112-R163, P221-L274, R325-P359, N372-Y425	BLIMPS_BLOCKS
					Met Apo-repressor, MetJ. PF01340: R390-N414	BLIMPS_PFAM
					TUBULIN CHAIN GTP BINDING MICROTUBULES PD000097: M1-Q423	BLAST_PRODOM
					TUBULIN SUBUNITS ALPHA, BETA, AND GAMMA DM00062 S18457|154-433: I156-G434 P41387|155-434: I156-E431 P52275|155-434: I156-E431 P08841|161-440: I156-E431	BLAST_DOMO
					Tubulin subunits alpha, beta, and gamma signature G141-G147	MOTIFS
					Tubulin-beta mRNA autoregulation signal M1-V5	MOTIFS
8	7612559CD1	198	S17 S19 S38 S144 S177 T63 T127 T155 Y114		Intermediate filament proteins: Q92-Y139	HMMER_PFAM
					Intermediate filaments proteins BL00226: Y80-R110, I121-K167	BLIMPS_BLOCKS
					Intermediate filaments signature if.prf: N133-Q183	PROFILESCAN
					INTERMEDIATE FILAMENTS DM00061|P23729|64-428: D94-D178	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	4940751CD1	139	S23 S52 S59 S94 S130 T31 T51 T76		Cofilin/tropomyosin-type actin-binding D12-R139	HMMER_PFAM
					Actin-depolymerizing proteins BL00325: G7-F38, D79-T124	BLIMPS_BLOCKS
					Cofilin/destrin family signature PR00006: D64-R84, F86-N107, Q108-T124	BLIMPS_PRINTS
					ACTIN-BINDING PROTEIN FACTOR CYTOSKELETON DEPOLYMERIZING COFILIN NUCLEAR PHOSPHORYLATION PD002129: N11-R137	BLAST_PRODOM
					ACTIN-DEPOLYMERIZING PROTEINS DM01110 P30175|4-138: S6-A138 P37167|1-132: S6-R139 Q03048|2-136: A4-R137 P54706|1-134: S6-I129	BLAST_DOMO
10	7946761CD1	736	S16 S41 S83 S91 S106 S349 S483 S630 T47 T110 T344 T388 T446 T458 T481 T593	N10	Signal_cleavage: M1-S37	SPSCAN
					Ank repeat: D160-R261	HMMER_PFAM
					Domain present in ZO-1 ankyrin receptors PF00791F: D578-I602 PF00791A: S218-A272 PF00791B: A168-D222 PF00791C: A182-G220 PF00791E: R381-P433	BLIMPS_PFAM
					PF00023A Ankyrin repeat protein domain L234-L249	BLIMPS_PFAM
					PD00078B Ankyrin repeat domain D227-Q239	BLIMPS_PRODOM
					PROTEIN K07D4.2 F42H11.2 PD155656: L598-I732	BLAST_PRODOM
					ANKYRIN REPEAT DM00014|A55575|160-206: L217-R259	BLAST_DOMO
					Cell attachment sequence R297-D299	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	3288747CD1	529	S31 S44 S145 S210 S249 S317 S362 S366 S422 S522 T6 T162 T230 T231 T248 T346 T433 Y247 Y325	N110 N459 N512	Signal_cleavage: M1-S29	SPSCAN
					Intermediate filament proteins: Q131-R444	HMMER_PFAM
					Transmembrane domain: F73-C95 N-terminus non-cystolic	TMAP
					Intermediate filaments proteins BL00226: Q131-S145, A232-Q279, D298-K328, L399-M445	BLIMPS_BLOCKS
					Intermediate filaments signature if.prf: A411-G469	PROFILESCAN
					FILAMENT INTERMEDIATE REPEAT HEPTAD PATTERN COILED COIL KERATIN PROTEIN TYPE PD000194: A130-R444, V107-R444	BLAST_PRODOM
					INTERMEDIATE FILAMENTS DM00061 A57398|126-498: L98-G467 P48666|125-497: L98-G467 P02538|125-497: L98-G467 I61768|126-498: L98-G467	BLAST_DOMO
					Leucine zipper pattern L183-L204, L389-L410	MOTIFS
					Cell attachment sequence R384-D386	MOTIFS
					Putative AMP-binding domain signature V513-R524	MOTIFS
					Intermediate filaments signature I431-E439	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	8200016CD1	1367	S211 S363 S378 S401 S411 S484 S497 S523 S626 S653 S677 S726 S784 S894 S984 S996 S1045 S1105 S1128 S1194 S1259 S1304 S1313 T200 T412 T430 T537 T567 T647 T751 T872 T1000 T1023	N81 N645 N818 N1067 N1225	Signal Peptide: M152-G181	HMMER
					Ank repeat: E243-C308, S114-V179	HMMER_PFAM
					Myosin head (motor domain): N425-G844, K866-K1155	HMMER_PFAM
					Transmembrane domain: W149-Q177, S607-L625, K668-D696 N-terminus cytosolic	TMAP
					Myosin heavy chain signature PR00193: Y453-Y472, P512-T537, C556-F583, D795-K823, A850-S878	BLIMPS_PRINTS
					Domain present in ZO-1 ankyrin receptor PF00791A: D136-E190 PF00791B: L248-N302 PF00791E: L462-C514	BLIMPS_PFAM
					PF00023A: Ankyrin repeat proteins L281-L296	BLIMPS_PFAM
					MYOSIN CHAIN HEAVY ATPBINDING ACTIN-BINDING PROTEIN COILED COIL MUSCLE MULTIGENE PD000355: I557-K1009, D426-T735, K1019-K1155, Q1191-T1262	BLAST_PRODOM
					MYOSIN HEAD DM00142 S38572|1-751: D426-E1031, I1004-Q1197 P08799|76-823: D426-F968, R1029-Q1187 P34092|1-727: K418-L995, L1030-I1195 S54307|136-1019: P421-L967, K997-Q1191, F1120-Q1187, K467-R550, E1350-R1360, K413-Y480	BLAST_DOMO
					Cell attachment sequence R1032-D1034	MOTIFS
					ATP/GTP-binding site motif A (P-loop) G519-S526	MOTIFS
					Myc-type, 'helix-loop-helix' dimerization domain signature E662-S677	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	3291962CD1	929	S123 S150 S253 S288 S292 S432 S449 S565 S703 S745 S748 T190 T271 T384 T472 T473 T507 T527 T727 T911 Y537 Y767	N73 N121 N520 N579 N743	TPR Domain: A43-N110, A6-K39	HMMER_PFAM
					PR00308B Type I Antifreeze protein domain A822-H833	BLIMPS_PRINTS
					PROTEIN CRO1 SHE4 RNG3 F30H5.1 CHROMOSOME III PD025764: L510-S745, Y715-L880, L486-C659, N350-S403	BLAST_PRODOM
					HYPOTHETICAL 107.4 KD PROTEIN F30H5.1 IN CHROMOSOME III PD146998: E115-D496, K689-A708	BLAST_PRODOM
					TPR REPEAT DM00408 P53041 | 24-181: A6-K127 P33313 | 79-231: A6-E126 S55383 | 397-559: E3-E126	BLAST_DOMO
14	1234259CD1	530	S3 S52 S84 S147 S180 S187 S217 S342 T32 T81 T93 T262 T308	N61 N209 N223 N277 N347	PUTATIVE TRANSCRIPTION FACTOR PD184883: D69-Q521	BLAST_PRODOM
15	1440608CD1	821	S70 S120 S138 S160 S164 S257 S266 S407 S417 S480 S481 S486 S488 S506 S523 S549 S588 S592 S668 T69 T90 T133 T137 T206 T239 T251 T744 T793 T805		Ank repeat: T695-A727, N622-R655, D728-N761, E762-Q794	HMMER_PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
16	3413610CD1	1003	S28 S151 S164 S204 S455 S662 S728 S775 S809 S844 S852 S997 T12 T100 T158 T498 T724 T750 T808 T860 T974 Y699 Y756	N33 N190 N357 N376 N585	BTB/POZ domain: R805-I921	HMMER_PFAM
					Ank repeat: Q502-V534, Y586-M618, R548-E580	HMMER_PFAM
					Predicted transmembrane segments: Q163-M191 E546-L567	TMAP
					Histone H2A signature PR00620: L199-V221, R228-S243	BLIMPS_PRINTS
					ANKYRIN PD144464: V10-G269, L392-Q502, T347-L381, D380-Q390, E289-A303, PD119546: L614-L800	BLAST_PRODOM
					Cell attachment sequence R513-D515	MOTIFS
17	3276394CD1	888	S22 S43 S58 S108 S156 S252 S304 S679 S712 S727 S800 S844 S883 T295 T299 T457 T490 T524 T691 T706 T728 Y172	N92 N106 N312 N510 N702 N824	SH3 domain: P137-I191, S448-V504, E832-I888, S199-N257	HMMER_PFAM
					Zinc finger, C3HC4 type (RING finger): C12-C52	HMMER_PFAM
					Src homology 3 (SH3) domain proteins profile BL50002: A141-D159, T490-P503	BLIMPS_BLOCKS
					PLENTY OF SH3S ZINCFINGER PD133543: P503-S839, PD086682: E255-N396 PD058054: T54-C138	BLAST_PRODOM
					Zinc finger, C3HC4 type (RING finger), signature C28-L37	MOTIFS
18	7602049CD1	283	S44 S83 S221 S227 S261 T61 T134 T235 T257	N75	signal_cleavage: M1-A38	SPSCAN
					PDZ domain (Also known as DHR or GLGF): S4-S83	HMMER_PFAM
					ENIGMA; LIM; RIL; DM03985 A55050|1-270: S2-P104, H151-G178, S205-E253 P52944|1-247: V5-S251, P50479|1-242: M1-K82	BLAST_DOMO

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
19/5566074CB1/1830	1-469, 13-158, 136-665, 157-425, 157-469, 170-506, 176-665, 237-691, 323-736, 336-836, 410-939, 522-984, 583-1134, 611-832, 660-930, 731-1317, 738-876, 836-1112, 846-1020, 874-1032, 922-1187, 938-1230, 960-1111, 960-1266, 960-1449, 960-1517, 960-1576, 960-1602, 1020-1146, 1033-1453, 1083-1576, 1104-1267, 1109-1407, 1190-1526, 1213-1567, 1239-1830, 1356-1821, 1366-1826, 1516-1797, 1537-1830, 1554-1830, 1648-1830, 1699-1830, 1739-1830
20/5679814CB1/2795	1-44, 1-67, 1-74, 1-82, 1-122, 1-246, 1-269, 1-309, 1-557, 17-611, 99-441, 145-698, 164-236, 164-244, 164-268, 164-269, 164-411, 189-709, 217-740, 329-599, 329-799, 351-417, 351-427, 353-953, 354-984, 411-655, 411-872, 424-476, 466-720, 466-1128, 474-958, 482-1135, 554-1165, 590-872, 593-853, 593-989, 595-871, 623-898, 628-1093, 649-1310, 674-1246, 702-1248, 710-1307, 730-1176, 746-1565, 746-1587, 779-1358, 792-1444, 802-1055, 813-1437, 816-929, 818-1333, 843-1474, 849-1350, 857-1465, 902-1386, 922-1184, 922-1297, 929-1512, 938-1529, 959-1500, 961-1486, 981-1572, 982-1534, 988-1082, 988-1257, 989-1120, 1000-1522, 1028-1647, 1037-1666, 1083-1775, 1108-1752, 1123-1584, 1127-1509, 1185-1766, 1196-1881, 1208-1504, 1215-1851, 1217-1375, 1217-1872, 1232-1911, 1244-1822, 1270-1872, 1286-1879, 1342-1872, 1368-1552, 1368-1860, 1373-1899, 1373-1966, 1466-1911, 1471-1765, 1471-2007, 1510-1649, 1510-1773, 1516-2052, 1555-2048, 1573-1865, 1581-1831, 1581-2032, 1597-2182, 1608-2218, 1622-2189, 1650-2271, 1670-1943, 1670-1954, 1750-1876, 1752-2269, 1759-2340, 1785-2475, 1804-2290, 1820-2271, 1825-2433, 1846-2129, 1852-2397,
	1876-1999, 1878-2086, 1878-2257, 1886-2175, 1886-2371, 1904-2340, 1913-2340, 1918-2562, 1921-2404, 1927-1999, 1933-1999, 1937-2193, 1938-2597, 1940-2439, 1948-2599, 1953-2230, 1987-2531, 1991-2208, 1992-2399, 2000-2339, 2003-2340, 2004-2166, 2006-2384, 2006-2506, 2030-2311, 2044-2357, 2087-2788, 2101-2571, 2112-2364, 2148-2795, 2188-2463, 2193-2783, 2232-2778, 2254-2456, 2296-2577, 2666-2723
21/7472735CB1/4436	1-216, 2-341, 22-670, 32-507, 32-597, 166-513, 429-699, 429-1012, 488-700, 682-1290, 734-1106, 804-1089, 859-1244, 896-1573, 1063-1665, 1245-1519, 1312-1766, 1457-1697, 1457-1942, 1631-2043, 1631-2141, 1631-2256, 1882-2117, 1895-2534, 1910-2182, 1937-2224, 1937-2353, 1937-2366, 1937-2407, 1939-2540, 1995-2632, 2062-2587, 2223-2719, 2292-2573, 2330-2625, 2333-2948, 2335-2649, 2407-2944, 2408-3051, 2416-2937, 2472-2666, 2489-2726, 2489-2975, 2489-3020, 2489-3091, 2489-3182, 2490-3162, 2534-3168, 2540-3028, 2542-3157, 2564-2716, 2565-2703, 2575-2797, 2576-2855, 2578-2874, 2578-3066, 2580-3144, 2599-2836, 2607-2822, 2607-2832, 2654-3173, 2666-3159, 2675-3321, 2686-3169, 2725-3382, 2744-3456, 2765-2989, 2765-3218, 2774-3475, 2786-3395, 2797-3315, 2824-3305, 2834-3264, 2836-3295, 2979-3510, 2988-3307, 3051-3327, 3085-3776, 3101-3828, 3113-3739, 3116-3384, 3117-3652, 3124-3306, 3136-3599, 3145-3596, 3148-3662, 3148-3815, 3161-3754, 3163-3429, 3165-3436, 3203-3482, 3240-3786, 3241-3860, 3257-3735, 3279-3522, 3304-3601, 3307-3546, 3307-3548, 3379-3666, 3394-3689, 3418-3681, 3511-3761, 3517-3760,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
21/7472735CB1/4436 (continued)	3542-3739, 3560-3813, 3587-3825, 3624-3851, 3644-3949, 3644-4214, 3645-4274, 3692-4420, 3748-3953, 3762-4053, 3774-4057, 3775-4069, 3790-4385, 3810-4029, 3817-4084, 3842-4392, 3871-4129, 3929-4133, 3942-4177, 4258-4436
22/7131221CB1/2040	1-308, 15-308, 15-647, 21-400, 29-693, 84-358, 103-731, 163-457, 163-767, 251-1005, 269-803, 452-651, 582-1371, 604-1119, 727-803, 802-1359, 812-1211, 812-1368, 827-1220, 829-1299, 834-1342, 869-1171, 869-1276, 893-1343, 901-1474, 917-1579, 918-1112, 936-1523, 962-1230, 962-1401, 963-1580, 1021-1173, 1033-1748, 1035-1061, 1061-1396, 1065-1345, 1065-1687, 1066-1779, 1086-1760, 1171-1400, 1185-1761, 1193-1753, 1213-1796, 1213-1802, 1234-1787, 1242-1805, 1265-1781, 1267-1526, 1267-1821, 1270-1821, 1327-1576, 1338-1631, 1338-1805, 1338-1821, 1418-1595, 1460-1821, 1608-2032, 1616-2040, 1643-1823, 1782-2020, 1802-2020
23/7480551CB1/2067	1-603, 52-636, 54-592, 60-697, 63-605, 66-536, 70-692, 79-638, 231-667, 296-551, 296-774, 316-632, 316-792, 461-785, 467-975, 473-692, 559-1075, 729-1060, 900-991, 944-1211, 944-1482, 1012-1403, 1139-1697, 1347-1696, 1567-2067
24/3315870CB1/1640	1-457, 34-223, 40-127, 42-223, 42-253, 56-305, 62-292, 66-364, 112-395, 201-321, 223-499, 357-623, 357-813, 365-1021, 462-508, 481-753, 488-753, 685-746, 716-1080, 784-1174, 878-1253, 894-1080, 898-1163, 898-1164, 901-1108, 901-1296, 917-1130, 917-1264, 931-1476, 961-1236, 961-1296, 964-1201, 1020-1474, 1060-1474, 1072-1533, 1082-1455, 1145-1479, 1176-1474, 1315-1640, 1318-1634, 1409-1478
25/7484690CB1/1497	1-498, 1-499, 1-500, 1-523, 1-542, 1-543, 1-544, 1-550, 1-571, 1-573, 1-575, 1-586, 1-587, 1-588, 1-593, 1-599, 1-618, 1-619, 1-631, 1-632, 1-690, 1-763, 1-1134, 5-582, 5-586, 21-678, 27-547, 28-650, 32-588, 58-141, 59-587, 59-708, 64-847, 68-676, 79-836, 103-654, 103-655, 103-847, 110-847, 111-714, 116-847, 118-714, 125-630, 125-800, 132-172, 139-740, 160-680, 160-847, 168-710, 169-215, 169-652, 172-663, 172-755, 177-648, 177-847, 187-632, 188-813, 193-808, 193-847, 202-847, 214-847, 227-847, 242-828, 258-715, 258-721, 258-731, 258-740, 258-759, 258-765, 258-772, 258-817, 258-850, 265-868, 281-481, 281-712, 281-730, 281-740, 281-771, 281-794, 281-851, 281-852, 281-867, 281-886, 281-975, 283-652, 283-760, 283-976, 283-982, 283-996, 283-1061, 283-1088, 285-1308, 287-827, 287-861, 298-815, 298-861, 306-894, 333-907, 342-941, 349-908, 351-812, 351-968, 353-957, 357-890, 359-928, 364-909, 374-929, 376-881, 384-851, 387-1201, 388-986, 388-993, 392-973, 406-990, 407-896, 414-1077, 424-909, 424-986, 424-1046, 424-1167, 425-993, 426-959, 429-977,
	430-986, 431-986, 437-932, 448-903, 457-993, 458-1073, 464-1040, 468-1067, 474-1094, 477-1122, 481-987, 484-1111, 485-1027, 488-986, 504-1000, 510-941, 512-986, 512-993, 512-1055, 512-1106, 517-1105, 519-1100, 520-993, 528-1094, 531-1035, 531-1118, 532-1118, 536-1065, 538-1090, 540-1118, 540-1130, 541-1130, 544-1111, 547-1130, 554-1196, 563-1040, 566-1216, 570-1092, 573-1154, 575-1084, 576-1094, 581-1103, 587-1232, 590-1112, 594-1105, 599-1106, 633-1112, 649-1183, 702-1477, 742-1406, 782-1255, 787-1497, 808-1312, 812-1497, 824-1497, 846-1497, 888-1437, 936-1480

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
26/7612559CB1/2065	1-283, 1-503, 5-486, 21-372, 41-292, 87-732, 146-376, 146-599, 189-786, 295-365, 349-868, 364-814, 364-909, 371-972, 374-630, 385-553, 393-582, 421-950, 434-1024, 470-993, 481-1024, 529-1110, 536-1069, 536-1174, 576-1054, 593-1166, 600-1287, 615-1247, 638-1312, 681-939, 684-973, 688-918, 688-1155, 700-979, 700-998, 736-984, 772-946, 802-1064, 937-1206, 957-1202, 976-1441, 1009-1255, 1058-1295, 1058-1660, 1083-1321, 1105-1379, 1140-1441, 1141-1420, 1148-1416, 1148-1424, 1155-1401, 1160-1405, 1172-1392, 1172-1776, 1180-1441, 1191-1409, 1192-1806, 1219-1470, 1219-1481, 1223-1863, 1237-1909, 1237-1917, 1245-1434, 1250-1915, 1259-1559, 1259-1560, 1269-1747, 1277-1930, 1312-1525, 1326-2065, 1334-1513, 1344-1900, 1347-1919, 1356-1904, 1372-1909, 1374-1637, 1377-1638, 1378-1686, 1378-1909, 1380-1931, 1402-1870, 1430-1899, 1430-1912, 1432-1914, 1446-1918, 1446-1920, 1484-1914, 1484-1920, 1487-1919, 1503-1919, 1523-1920, 1528-1764, 1528-1907, 1528-1920, 1533-1894, 1540-1931, 1555-1920, 1573-1931, 1574-1703, 1593-1919, 1595-1919, 1605-1920, 1605-1929, 1606-1919, 1607-1850, 1608-1931, 1615-1920, 1616-1862, 1617-1919, 1626-1920, 1634-1908, 1637-1931, 1641-1920, 1646-1920, 1649-1926, 1654-1904, 1682-1919, 1683-1920, 1697-1920, 1725-1920, 1731-1920, 1731-1929, 1734-1918, 1752-1921, 1798-1920
27/4940751CB1/762	1-554, 27-305, 159-762
28/7946761CB1/2211	1-2211, 1556-2116, 1557-1994, 1557-2115, 1558-2116, 1736-2187, 1908-2211, 2081-2211, 2082-2211
29/3288747CB1/1634	1-253, 11-602, 11-685, 11-707, 48-1019, 48-1634, 435-1046, 496-864, 496-916, 714-999, 836-1207, 1061-1569, 1109-1431, 1157-1207
30/8200016CB1/4706	1-698, 1-743, 1-860, 6-860, 25-860, 57-860, 62-860, 111-860, 612-682, 655-860, 683-826, 701-1309, 701-1402, 861-1578, 1000-1578, 1025-1578, 1294-1742, 1324-1542, 1565-1889, 1669-2243, 2011-2550, 2190-2829, 2394-3029, 2715-3323, 2720-3324, 3100-3313, 3260-3770, 3589-3802, 3590-4095, 3660-4479, 3924-4303, 3930-4440, 3958-4557, 3958-4706, 4044-4500

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
31/3291962CB1/3029	1-233, 1-430, 1-582, 28-288, 28-453, 28-454, 28-550, 28-595, 29-557, 30-626, 64-589, 277-954, 307-1012, 309-601, 363-914, 520-1115, 520-1144, 653-1287, 702-1241, 708-1330, 787-1470, 807-1257, 807-1470, 816-1445, 821-1213, 828-1507, 855-1492, 897-1487, 956-1239, 956-1478, 956-1487, 956-1514, 977-1627, 1007-1509, 1023-1653, 1023-1729, 1024-1584, 1026-1630, 1081-1645, 1118-1726, 1180-1639, 1188-1726, 1188-1779, 1215-1859, 1258-1948, 1258-1949, 1281-1485, 1310-1676, 1320-1825, 1356-1960, 1360-1930, 1406-1829, 1419-1846, 1516-1784, 1520-1770, 1572-2223, 1582-2257, 1595-1985, 1631-2189, 1704-2291, 1713-2359, 1725-1953, 1727-2307, 1781-2327, 1820-2225, 1852-2359, 1857-2455, 1928-2572, 1964-2329, 1974-2578, 1996-2456, 2006-2525, 2039-2649, 2055-2309, 2074-2642, 2082-2678, 2135-2737, 2141-2521, 2168-2779, 2193-2566, 2229-2518, 2259-2914, 2290-2780, 2314-2699, 2345-2727, 2390-2584, 2410-2972, 2411-2998, 2506-2975, 2529-3029, 2560-3029, 2751-2908
32/1234259CB1/2074	1-151, 1-197, 1-219, 1-244, 1-264, 1-514, 17-546, 26-283, 26-646, 26-652, 26-653, 26-701, 26-844, 27-502, 35-317, 43-274, 43-527, 46-255, 69-340, 75-422, 154-358, 154-368, 299-900, 306-831, 306-881, 316-795, 386-1199, 392-1175, 401-1050, 406-1191, 409-952, 445-1047, 470-1325, 528-1178, 530-977, 534-1319, 555-832, 564-1025, 568-1166, 586-1412, 589-1316, 595-1066, 631-1057, 631-1339, 636-1412, 647-1466, 686-1221, 710-1414, 719-1114, 721-1516, 734-837, 740-978, 742-1469, 757-1558, 779-1412, 809-1493, 850-1510, 859-998, 894-1637, 897-1056, 901-1652, 905-1665, 914-1556, 915-1542, 919-1545, 925-1182, 940-1183, 953-1259, 965-1666, 974-1131, 980-1508, 992-1520, 1008-1350, 1016-1816, 1034-1288, 1034-1601, 1044-1330, 1049-1680, 1052-1513, 1061-1294, 1066-1725, 1079-1626, 1086-1347, 1093-1297, 1094-1255, 1094-1709, 1095-1867, 1113-1630, 1144-1618, 1152-1302, 1155-1647, 1175-1598, 1181-1507, 1271-2074, 1303-2000, 1315-1338, 1348-1907, 1401-2043, 1513-2056, 1515-2043, 1523-1916, 1580-1701, 1653-1910, 1657-1805, 1672-2072, 1675-2074
33/1440608CB1/2710	1-34, 1-1290, 1044-1463, 1045-1243, 1045-1412, 1045-1414, 1045-1429, 1045-1531, 1045-1612, 1048-1528, 1063-1429, 1068-1556, 1079-1466, 1084-1265, 1084-1536, 1094-1334, 1158-1803, 1202-1435, 1263-1879, 1281-1529, 1431-1775, 1479-1506, 1529-1783, 1541-1984, 1605-1850, 1605-2100, 1634-2008, 1693-2002, 1693-2194, 1728-1966, 1767-2026, 1812-2092, 1817-2118, 1854-2482, 1866-2112, 1886-2180, 1895-2104, 1899-2146, 1908-2149, 1908-2158, 1908-2339, 1936-2182, 1943-2669, 1969-2182, 2062-2309, 2122-2390, 2218-2710, 2372-2572, 2372-2598, 2372-2710, 2379-2634, 2434-2710, 2464-2710, 2546-2710, 2632-2710

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
34/3413610CB1/3527	1-1135, 968-1135, 1014-1135, 1014-1436, 1014-1441, 1062-1135, 1276-1909, 1412-1480, 1427-1963, 1427-2032, 1758-2191, 1758-2202, 1758-2207, 1758-2243, 1758-2263, 1758-2282, 1758-2300, 1758-2302, 1758-2379, 1760-1972, 1838-2501, 1864-2043, 1896-2131, 1896-2142, 1896-2220, 1896-2407, 1896-2441, 1900-2495, 1934-2578, 1943-2453, 1979-2265, 2093-2624, 2096-2349, 2096-2710, 2096-2750, 2181-2806, 2203-2620, 2212-2885, 2275-2764, 2324-2868, 2326-2415, 2355-2960, 2402-2683, 2402-2857, 2402-2909, 2413-3011, 2419-3078, 2423-2929, 2427-3123, 2428-2985, 2454-3032, 2457-3118, 2465-3037, 2494-3067, 2506-3044, 2507-2947, 2511-2762, 2511-3181, 2514-3046, 2541-3112, 2556-2811, 2566-2881, 2569-3228, 2576-3162, 2579-2831, 2585-3240, 2586-2892, 2586-3213, 2593-3224, 2680-3237, 2696-3284, 2699-3374, 2701-3134, 2710-3234, 2721-3347, 2753-3298, 2753-3416, 2759-3273, 2793-3502, 2796-3408, 2796-3501, 2796-3502, 2816-3065, 2823-3338, 2823-3358, 2834-3334, 2835-3077, 2841-3365, 2841-3405, 2851-3448, 2866-3405, 2872-3484, 2881-3448, 2884-3245, 2884-3258, 2884-3292, 2884-3309, 2884-3312, 2884-3358, 2888-3192,
	2888-3246, 2888-3252, 2888-3352, 2888-3358, 2889-3360, 2890-3355, 2893-3527, 2901-3105, 2913-3424, 2937-3521, 2944-3448, 2960-3360, 2960-3448, 2962-3527, 2965-3527, 2969-3203, 2974-3527, 2979-3527, 2986-3080, 2987-3527, 3003-3527, 3010-3510, 3010-3513, 3010-3517, 3012-3517, 3020-3527, 3026-3500, 3027-3527, 3033-3250, 3036-3523, 3039-3527, 3042-3527, 3055-3527, 3057-3362, 3057-3370, 3069-3405, 3076-3527, 3082-3527, 3095-3465, 3106-3471, 3110-3527, 3140-3327, 3157-3495, 3400-3525, 3412-3525
35/3276394CB1/3251	1-594, 21-665, 26-624, 26-644, 28-586, 43-675, 104-268, 104-366, 104-378, 104-593, 107-351, 137-593, 165-593, 170-593, 188-593, 217-593, 272-593, 276-593, 295-593, 341-850, 349-593, 498-999, 505-808, 551-1154, 554-1105, 609-1084, 742-1302, 743-1359, 823-1398, 879-1529, 883-1412, 891-1169, 947-1382, 1034-1502, 1058-1714, 1063-1622, 1094-1697, 1097-1603, 1120-1706, 1168-1823, 1172-1533, 1195-1766, 1275-1887, 1295-1940, 1313-1937, 1364-1937, 1389-1623, 1397-1984, 1499-1989, 1573-2076, 1598-2110, 1598-2175, 1604-2297, 1662-2324, 1677-2323, 1681-1909, 1711-2060, 1727-2312, 1746-1981, 1746-2017, 1746-2033, 1798-2111, 1798-2350, 1806-2351, 1857-2254, 1859-2298, 1863-2262, 1863-2350, 1868-2307, 1870-1957, 1873-2093, 1893-2350, 1898-2342, 1936-2336, 1977-2462, 1994-2336, 2008-2339, 2020-2315, 2041-2288, 2046-2310, 2091-2350, 2144-2323, 2144-2335, 2147-2320, 2162-2278, 2189-2776, 2200-2337, 2265-2845, 2268-2905, 2371-2511, 2576-3211, 2589-3251, 2611-2860, 2728-3036, 2742-2899, 2742-2934, 2746-3203, 2777-3151, 2779-3105

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
36/7602049CB1/1600	1-603, 13-649, 24-453, 41-378, 42-631, 80-363, 102-434, 113-357, 113-511, 113-574, 113-600, 113-615, 113-647, 113-687, 113-721, 128-403, 180-384, 232-472, 294-698, 294-812, 309-855, 328-561, 328-955, 385-818, 390-805, 407-786, 423-708, 450-1016, 451-762, 455-1051, 463-661, 487-1140, 501-795, 509-1101, 510-1223, 516-1197, 521-742, 521-1006, 541-1167, 556-1055, 626-849, 626-1114, 637-939, 642-890, 666-870, 667-1263, 670-855, 680-967, 686-992, 709-965, 795-976, 810-1065, 816-993, 823-1078, 833-1120, 879-1060, 896-1175, 926-1566, 927-1555, 945-1187, 969-1209, 969-1515, 974-1574, 983-1580, 987-1264, 1005-1579, 1054-1331, 1139-1281, 1155-1453, 1195-1583, 1213-1568, 1213-1572, 1213-1599, 1248-1538, 1250-1543, 1293-1585, 1350-1600, 1383-1556, 1383-1600, 1452-1600

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
19	5566074CB1	BRACNOK02
20	5679814CB1	OVARNOT09
21	7472735CB1	BRALNOT01
22	7131221CB1	MUSCNOT11
23	7480551CB1	BRSTNOT16
24	3315870CB1	BRSTNOT35
25	7484690CB1	TESTTUE02
26	7612559CB1	ADRENOT07
27	4940751CB1	BRAIFEN03
28	7946761CB1	LIVRFEE02
29	3288747CB1	LNODNON02
30	8200016CB1	BRAIFER06
31	3291962CB1	BONRFET01
32	1234259CB1	PROSNOT16
33	1440608CB1	SINTNOT02
34	3413610CB1	PROTDNV09
35	3276394CB1	CONFNOT07
36	7602049CB1	MUSCNOT01

Table 6

Library	Vector	Library Description
ADRENOT07	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
BONRFET01	pINCY	Library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.
BRACNOK02	PSPORT1	This amplified and normalized library was constructed using RNA isolated from posterior cingulate tissue removed from an 85-year-old Caucasian female who died from myocardial infarction and retroperitoneal hemorrhage. Pathology indicated atherosclerosis, moderate to severe, involving the circle of Willis, middle cerebral, basilar and vertebral arteries; infarction, remote, left dentate nucleus; and amyloid plaque deposition consistent with age. There was mild to moderate leptomeningeal fibrosis, especially over the convexity of the frontal lobe. There was mild generalized atrophy involving all lobes. The white matter was mildly thinned. Cortical thickness in the temporal lobes, both maximal and minimal, was slightly reduced. The substantia nigra pars compacta appeared mildly depigmented. Patient history included COPD, hypertension, and recurrent deep venous thrombosis. 6.4 million independent clones from this amplified library were normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791.
BRAIFEN03	pINCY	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRALNOT01	pINCY	Library was constructed using RNA isolated from thalamus tissue removed from a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.
BRSTNOT16	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 59-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive lobular carcinoma with extension into ducts. Patient history included liver cirrhosis, esophageal ulcer, hyperlipidemia, and neuropathy.

Table 6

Library	Vector	Library Description
BRSTNOT35	pINCY	Library was constructed using RNA isolated from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated normal breast parenchyma, bilaterally. The patient presented with hypertrophy of breast and headache. Patient history included obesity, lumbago, glaucoma, and alcohol abuse. Family history included cataract,osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes.
CONFNOT07	pINCY	Library was constructed using RNA isolated from abdominal adipose tissue removed from a 68-year-old Caucasian female during open cholecystectomy and ventral hernia repair. Patient history included morbid obesity, cholelithiasis, ventral hernia, mitral valve prolapse, hypothyroidism, myocardial infarction, and uterine cancer.
LIVRFEE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from liver tissue removed from a Caucasian male fetus who died from fetal demise. Serologies were negative.
LNODNON02	pINCY	This normalized lymph node tissue library was constructed from .56 million independent clones from a lymph node tissue library. Starting RNA was made from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Serologies were negative. Patient history included bronchitis. Patient medications included Dopamine, Dobutamine, Vancomycin, Vasopressin, Proventil, and Atarax. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9932 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
MUSCNOT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937209), using RNA isolated from the skeletal muscle tissue of a patient with malignant hyperthermia.
MUSCNOT11	pINCY	The library was constructed using RNA isolated from diseased arm muscle tissue removed from a 74-year-old Caucasian female who died from respiratory arrest due to amyotrophic lateral sclerosis (ALS). Patient historyincluded amyotrophic lateral sclerosis, hypertension, arthritis, and alcohol use.
OVARNOT09	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis and squamous metaplasia of the cervix, and endometrium in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.

Table 6

Library	Vector	Library Description
ADRENOT07	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
BONRFET01	pINCY	Library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.
BRACNOK02	PSPORT1	This amplified and normalized library was constructed using RNA isolated from posterior cingulate tissue removed from an 85-year-old Caucasian female who died from myocardial infarction and retroperitoneal hemorrhage. Pathology indicated atherosclerosis, moderate to severe, involving the circle of Willis, middle cerebral, basilar and vertebral arteries; infarction, remote, left dentate nucleus; and amyloid plaque deposition consistent with age. There was mild to moderate leptomeningeal fibrosis, especially over the convexity of the frontal lobe. There was mild generalized atrophy involving all lobes. The white matter was mildly thinned. Cortical thickness in the temporal lobes, both maximal and minimal, was slightly reduced. The substantia nigra pars compacta appeared mildly depigmented. Patient history included COPD, hypertension, and recurrent deep venous thrombosis. 6.4 million independent clones from this amplified library were normalized in one round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791.
BRAIFEN03	pINCY	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRALNOT01	pINCY	Library was constructed using RNA isolated from thalamus tissue removed from a 35-year-old Caucasian male. No neuropathology was found. Patient history included dilated cardiomyopathy, congestive heart failure, and an enlarged spleen and liver.
BRSTNOT16	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 59-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive lobular carcinoma with extension into ducts. Patient history included liver cirrhosis, esophageal ulcer, hyperlipidemia, and neuropathy.

Table 6

Library	Vector	Library Description
BRSTNOT35	pINCY	Library was constructed using RNA isolated from breast tissue removed from a 46-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated normal breast parenchyma, bilaterally. The patient presented with hypertrophy of breast and headache. Patient history included obesity, lumbago, glaucoma, and alcohol abuse. Family history included cataract,osteoarthritis, uterine cancer, benign hypertension, hyperlipidemia, alcoholic cirrhosis of the liver, cerebrovascular disease, and type II diabetes.
CONFNOT07	pINCY	Library was constructed using RNA isolated from abdominal adipose tissue removed from a 68-year-old Caucasian female during open cholecystectomy and ventral hernia repair. Patient history included morbid obesity, cholelithiasis, ventral hernia, mitral valve prolapse, hypothyroidism, myocardial infarction, and uterine cancer.
LIVRFEE02	pINCY	This 5' biased random primed library was constructed using RNA isolated from liver tissue removed from a Caucasian male fetus who died from fetal demise. Serologies were negative.
LNODNON02	pINCY	This normalized lymph node tissue library was constructed from .56 million independent clones from a lymph node tissue library. Starting RNA was made from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Serologies were negative. Patient history included bronchitis. Patient medications included Dopamine, Dobutamine, Vancomycin, Vasopressin, Proventil, and Atarax. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9932 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
MUSCNOT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937209), using RNA isolated from the skeletal muscle tissue of a patient with malignant hyperthermia.
MUSCNOT11	pINCY	The library was constructed using RNA isolated from diseased arm muscle tissue removed from a 74-year-old Caucasian female who died from respiratory arrest due to amyotrophic lateral sclerosis (ALS). Patient historyincluded amyotrophic lateral sclerosis, hypertension, arthritis, and alcohol use.
OVARNOT09	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 28-year-old Caucasian female during a vaginal hysterectomy and removal of the fallopian tubes and ovaries. Pathology indicated multiple follicular cysts ranging in size from 0.4 to 1.5 cm in the right and left ovaries, chronic cervicitis and squamous metaplasia of the cervix, and endometrium in weakly proliferative phase. Family history included benign hypertension, hyperlipidemia, and atherosclerotic coronary artery disease.

Table 6

Library	Vector	Library Description
PROSNOT16	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
PROTDNV09	PCR2-TOPOTA	Library was constructed using pooled cDNA from 106 different donors. cDNA was generated using mRNA isolated from lung tissue removed from male Caucasian fetus (donor A) who died from fetal demise; from brain and small intestine tissue removed from a 23-week-old Caucasian male fetus (donor B) who died from premature birth; from brain tissue removed from a Caucasian male fetus (donor C) who was stillborn with a hypoplastic left heart at 23 weeks' gestation; from liver tumor tissue removed from a 72-year-old Caucasian male (donor D) during partial hepatectomy; from left frontal/parietal brain tumor tissue removed from a 2-year-old Caucasian female (donor E) during excision of cerebral meningeal lesion; from pleural tumor tissue removed from a 55-year-old Caucasian female (donor F) during complete pneumonectomy; from liver tissue removed from a pool of thirty-two, 18 to 24-week-old male and female fetuses (donor G) who died from spontaneous abortions; from kidney tissue removed from a pool of fifty-nine 20 to 33-week-old male and
		female fetuses (donor H) who died from spontaneous abortions; and from thymus tissue removed from a pool of nine 18 to 32-year-old males and females (donor I) who died from sudden death. For donors A, B, and C, serologies were negative. For donor B, family history included diabetes in the mother. For donor D, pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma of the right liver lobe. The patient presented with secondary malignant neoplasm of the liver. Patient history included benign hypertension, type I diabetes, hyperplasia of the prostate, malignant prostate neoplasm, and tobacco and alcohol abuse in remission. Previous surgeries included excision/destruction of a pancreas lesion (insulinoma), closed prostatic biopsy, transurethral prostatectomy, and excision of both testes. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included acute myocardial infarction and atherosclerotic coronary artery disease in the mother, and atherosclerotic coronary artery disease and type II diabetes in the father. For donor

Table 6

Library	Vector	Library Description
PROTDNV09 (continued)		E, pathology indicated primitive neuroectodermal tumor with advanced ganglionic differentiation. The lesion was only moderately cellular but was mitotically active with a high MIB-1 labelling index. Neuronal differentiation was widespread and advanced. Multinucleate and dysplastic-appearing forms were readily seen. The glial element was less prominent. Synaptophysin, GFAP, and S-100 were positive. The patient presented with malignant brain neoplasm and motor seizures. The patient was not taking any medications. Family history included benign hypertension in the grandparent(s). For donor F, pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, involving the parietal pleura. The patient presented with secondary malignant lung neoplasm and shortness of breath. Patient history included peptic ulcer disease, malignant uterine neoplasm, normal delivery, deficiency anemia, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy,
		hemorrhoidectomy, endoscopic excision of lung lesion, and incidental appendectomy. Patient medications included Megace, Pepcid and tamoxifen. Family history included atherosclerotic coronary artery disease and type II diabetes in the father; multiple sclerosis in the mother; and malignant breast neoplasm in the grandparent(s).
SINTNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from the small intestine of a 55-year-old Caucasian female, who died from a subarachnoid hemorrhage. Serologies were positive for cytomegalovirus (CMV). Previous surgeries included a hysterectomy.
TESTTUE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma forming a largely necrotic mass involving the entire testicle. Rare foci of residual testicle showed intralobular germ cell neoplasia and tumor was identified at the spermatic cord margin. The patient presented with backache. Patient history included tobacco use. Previous surgeries included a needle biopsy of testis. Patient medications included Colace and antacids.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less; Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6; Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less; Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less; Signal peptide hits: Score= 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18,

b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-6 and SEQ ID NO:10-18,

c) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:7,

d) a polypeptide comprising a naturally occuring amino acid sequence at least 98% identical to the amino acid sequence of SEQ ID NO:8,

e) a polypeptide comprising a naturally occuring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:9,

f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and

g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method of producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,

c) a polynucleotide complementary to a polynucleotide of a),

d) a polynucleotide complementary to a polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target

polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

19. A method for treating a disease or condition associated with decreased expression of functional CSAP, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional CSAP, comprising administering to a patient in need of such treatment a composition of

claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional CSAP, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target

polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

a) treating a biological sample containing nucleic acids with the test compound,

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

c) quantifying the amount of hybridization complex, and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of CSAP in a biological sample, the method comprising:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody,

b) a single chain antibody,

c) a Fab fragment,

d) a $F(ab')_2$ fragment, or

e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of CSAP in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of CSAP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibodies from said animal, and

c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

b) isolating antibody producing cells from the animal,

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,

d) culturing the hybridoma cells, and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and

b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

a) labeling the polynucleotides of the sample,

b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.

73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

80. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:25.

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

<110> INCYTE GENOMICS, INC.
LU, Dyung Aina M.
BAUGHN, Mariah R.
YAO, Monique G.
DING, Li
HONCHELL, Cynthia D.
YUE, Henry
TANG, Y. Tom
WARREN, Bridget A.
DUGGAN, Brendan M.
XU, Yuming
WALIA, Narinder K.
GRIFFIN, Jennifer A.
STEWART, Elizabeth A.
GANDHI, Ameena R.
KHAN, Farrah A.
THANGAVELU, Kavitha
ISON, Craig H.
AZIMZAI, Yalda
HAFALIA, April J.A.
GIETZEN, Kimberly J.
LAL, Preeti G.
SANJANWALA, Madhu M.
ELLIOTT, Vicki S.

<120> CYTOSKELETAL-ASSOCIATED PROTEINS

<130> PF-0878 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/260,085; 60/268,554; 60/269,111; 60/271,211
<151> 2001-01-04; 2001-02-13; 2001-02-14; 2001-02-23

<160> 36
<170> PERL Program

<210> 1
<211> 377
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5566074CD1

<400> 1
```
Met Tyr Thr Phe Val Val Arg Asp Glu Asn Ser Ser Val Tyr Ala
  1               5                  10                  15
Glu Val Ser Arg Leu Leu Leu Ala Thr Gly His Trp Lys Arg Leu
                 20                  25                  30
Arg Arg Asp Asn Pro Arg Phe Asn Leu Met Leu Gly Glu Arg Asn
                 35                  40                  45
Arg Leu Pro Phe Gly Arg Leu Gly His Glu Pro Gly Leu Val Gln
                 50                  55                  60
```

```
Leu Val Asn Tyr Tyr Arg Gly Ala Asp Lys Leu Cys Arg Lys Ala
                65                  70                  75
Ser Leu Val Lys Leu Ile Lys Thr Ser Pro Glu Leu Ala Glu Ser
                80                  85                  90
Cys Thr Trp Phe Pro Glu Ser Tyr Val Ile Tyr Pro Thr Asn Leu
                95                  100                 105
Lys Thr Pro Val Ala Pro Ala Gln Asn Gly Ile Gln Pro Pro Ile
                110                 115                 120
Ser Asn Ser Arg Thr Asp Glu Arg Glu Phe Phe Leu Ala Ser Tyr
                125                 130                 135
Asn Arg Lys Lys Glu Asp Gly Glu Gly Asn Val Trp Ile Ala Lys
                140                 145                 150
Ser Ser Ala Gly Ala Lys Gly Glu Gly Ile Leu Ile Ser Ser Glu
                155                 160                 165
Ala Ser Glu Leu Leu Asp Phe Ile Asp Asn Gln Gly Gln Val His
                170                 175                 180
Val Ile Gln Lys Tyr Leu Glu His Pro Leu Leu Leu Glu Pro Gly
                185                 190                 195
His Arg Lys Phe Asp Ile Arg Ser Trp Val Leu Val Asp His Gln
                200                 205                 210
Tyr Asn Ile Tyr Leu Tyr Arg Glu Gly Val Leu Arg Thr Ala Ser
                215                 220                 225
Glu Pro Tyr His Val Asp Asn Phe Gln Asp Lys Thr Cys His Leu
                230                 235                 240
Thr Asn His Cys Ile Gln Lys Glu Tyr Ser Lys Asn Tyr Gly Lys
                245                 250                 255
Tyr Glu Glu Gly Asn Glu Met Phe Phe Lys Glu Phe Asn Gln Tyr
                260                 265                 270
Leu Thr Ser Ala Leu Asn Ile Thr Leu Glu Ser Ser Ile Leu Leu
                275                 280                 285
Gln Ile Lys His Ile Ile Arg Asn Cys Leu Leu Ser Val Glu Pro
                290                 295                 300
Ala Ile Ser Thr Lys His Leu Pro Tyr Gln Ser Phe Gln Leu Phe
                305                 310                 315
Gly Phe Asp Phe Met Val Asp Glu Glu Leu Lys Val Trp Leu Ile
                320                 325                 330
Glu Val Asn Gly Ala Pro Ala Cys Ala Gln Lys Leu Tyr Ala Glu
                335                 340                 345
Leu Cys Gln Gly Ile Val Asp Ile Ala Ile Ser Ser Val Phe Pro
                350                 355                 360
Pro Pro Asp Val Glu Gln Pro Gln Thr Gln Pro Ala Ala Phe Ile
                365                 370                 375
Lys Leu
```

<210> 2
<211> 696
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5679814CD1

<400> 2
Met Lys Trp Leu Ile Asp Pro Leu Pro Val Asn Val Arg Val Ile

```
1               5                   10                  15
Val Ser Val Asn Val Glu Thr Cys Pro Pro Ala Trp Arg Leu Trp
                20                  25                  30
Pro Thr Leu His Leu Asp Pro Leu Ser Pro Lys Asp Ala Lys Ser
                35                  40                  45
Ile Ile Ile Ala Glu Cys His Ser Val Asp Ile Lys Leu Ser Lys
                50                  55                  60
Glu Gln Glu Lys Lys Leu Glu Arg His Cys Arg Ser Ala Thr Thr
                65                  70                  75
Cys Asn Ala Leu Tyr Val Thr Leu Phe Gly Lys Met Ile Ala Arg
                80                  85                  90
Ala Gly Arg Ala Gly Asn Leu Asp Lys Ile Leu His Gln Cys Phe
                95                  100                 105
Gln Cys Gln Asp Thr Leu Ser Leu Tyr Arg Leu Val Leu His Ser
                110                 115                 120
Ile Arg Glu Ser Met Ala Asn Asp Val Asp Lys Glu Leu Met Lys
                125                 130                 135
Gln Ile Leu Cys Leu Val Asn Val Ser His Asn Gly Val Ser Glu
                140                 145                 150
Ser Glu Leu Met Glu Leu Tyr Pro Glu Met Ser Trp Thr Phe Leu
                155                 160                 165
Thr Ser Leu Ile His Ser Leu Tyr Lys Met Cys Leu Leu Thr Tyr
                170                 175                 180
Gly Cys Gly Leu Leu Arg Phe Gln His Leu Gln Ala Trp Glu Thr
                185                 190                 195
Val Arg Leu Glu Tyr Leu Glu Gly Pro Thr Val Thr Ser Ser Tyr
                200                 205                 210
Arg Gln Lys Leu Ile Asn Tyr Phe Thr Leu Gln Leu Ser Gln Asp
                215                 220                 225
Arg Val Thr Trp Arg Ser Ala Asp Glu Leu Pro Trp Leu Phe Gln
                230                 235                 240
Gln Gln Gly Ser Lys Gln Lys Leu His Asp Cys Leu Leu Asn Leu
                245                 250                 255
Phe Val Ser Gln Asn Leu Tyr Lys Arg Gly His Phe Ala Glu Leu
                260                 265                 270
Leu Ser Tyr Trp Gln Phe Val Gly Lys Asp Lys Ser Ala Met Ala
                275                 280                 285
Thr Glu Tyr Phe Asp Ser Leu Lys Gln Tyr Glu Lys Asn Cys Glu
                290                 295                 300
Gly Glu Asp Asn Met Ser Cys Leu Ala Asp Leu Tyr Glu Thr Leu
                305                 310                 315
Gly Arg Phe Leu Lys Asp Leu Gly Leu Leu Ser Gln Ala Ile Val
                320                 325                 330
Pro Leu Gln Arg Ser Leu Glu Ile Arg Glu Thr Ala Leu Asp Pro
                335                 340                 345
Asp His Pro Arg Val Ala Gln Ser Leu His Gln Leu Ala Ser Val
                350                 355                 360
Tyr Val Gln Trp Lys Lys Phe Gly Asn Ala Glu Gln Leu Tyr Lys
                365                 370                 375
Gln Ala Leu Glu Ile Ser Glu Asn Ala Tyr Gly Ala Asp His Pro
                380                 385                 390
Tyr Thr Ala Arg Glu Leu Glu Ala Leu Ala Thr Leu Tyr Gln Lys
                395                 400                 405
Gln Asn Lys Tyr Glu Gln Ala Glu His Phe Arg Lys Lys Ser Phe
                410                 415                 420
Lys Ile His Gln Lys Ala Ile Lys Lys Lys Gly Asn Leu Tyr Gly
```

425 430 435
Phe Ala Leu Leu Arg Arg Arg Ala Leu Gln Leu Glu Glu Leu Thr
440 445 450
Leu Gly Lys Asp Thr Pro Asp Asn Ala Arg Thr Leu Asn Glu Leu
455 460 465
Gly Val Leu Tyr Tyr Leu Gln Asn Asn Leu Glu Thr Ala Asp Gln
470 475 480
Phe Leu Lys Arg Ser Leu Glu Met Arg Glu Arg Val Leu Gly Pro
485 490 495
Asp His Pro Asp Cys Ala Gln Ser Leu Asn Asn Leu Ala Ala Leu
500 505 510
Cys Asn Glu Lys Lys Gln Tyr Asp Lys Ala Glu Glu Leu Tyr Glu
515 520 525
Arg Ala Leu Asp Ile Arg Arg Arg Ala Leu Ala Pro Asp His Pro
530 535 540
Ser Leu Ala Tyr Thr Val Lys His Leu Ala Ile Leu Tyr Lys Lys
545 550 555
Met Gly Lys Leu Asp Lys Ala Val Pro Leu Tyr Glu Leu Ala Val
560 565 570
Glu Ile Arg Gln Lys Ser Phe Gly Pro Lys His Pro Ser Val Ala
575 580 585
Thr Ala Leu Val Asn Leu Ala Val Leu Tyr Ser Gln Met Lys Lys
590 595 600
His Val Glu Ala Leu Pro Leu Tyr Glu Arg Ala Leu Lys Ile Tyr
605 610 615
Glu Asp Ser Leu Gly Arg Met His Pro Arg Val Gly Glu Thr Leu
620 625 630
Lys Asn Leu Ala Val Leu Ser Tyr Glu Gly Gly Asp Phe Glu Lys
635 640 645
Ala Ala Glu Leu Tyr Lys Arg Ala Met Glu Ile Lys Glu Ala Glu
650 655 660
Thr Ser Leu Leu Gly Gly Lys Ala Pro Ser Arg His Ser Ser Ser
665 670 675
Gly Asp Thr Phe Ser Leu Lys Thr Ala His Ser Pro Asn Val Phe
680 685 690
Leu Gln Gln Gly Gln Arg
695

<210> 3
<211> 1050
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472735CD1

<400> 3
Met Ala Leu Tyr Asp Glu Asp Leu Leu Lys Asn Pro Phe Tyr Leu
1 5 10 15
Ala Leu Gln Lys Cys Arg Pro Asp Leu Cys Ser Lys Val Ala Gln
20 25 30
Ile His Gly Ile Val Leu Val Pro Cys Lys Gly Ser Leu Ser Ser
35 40 45
Ser Ile Gln Ser Thr Cys Gln Phe Glu Ser Tyr Ile Leu Ile Pro
50 55 60

Val Glu Glu His Phe Gln Thr Leu Asn Gly Lys Asp Val Phe Ile
65 70 75
Gln Gly Asn Arg Ile Lys Leu Gly Ala Gly Phe Ala Cys Leu Leu
80 85 90
Ser Val Pro Ile Leu Phe Glu Glu Thr Phe Tyr Asn Glu Lys Glu
95 100 105
Glu Ser Phe Ser Ile Leu Cys Ile Ala His Pro Leu Glu Lys Arg
110 115 120
Glu Ser Ser Glu Glu Pro Leu Ala Pro Ser Asp Pro Phe Ser Leu
125 130 135
Lys Thr Ile Glu Asp Val Arg Glu Phe Leu Gly Arg His Ser Glu
140 145 150
Arg Phe Asp Arg Asn Ile Ala Ser Phe His Arg Thr Phe Arg Glu
155 160 165
Cys Glu Arg Lys Ser Leu Arg His His Ile Asp Ser Ala Asn Ala
170 175 180
Leu Tyr Thr Lys Cys Leu Gln Gln Leu Leu Arg Asp Ser His Leu
185 190 195
Lys Met Leu Ala Lys Gln Glu Ala Gln Met Asn Leu Met Lys Gln
200 205 210
Ala Val Glu Ile Tyr Val His His Glu Ile Tyr Asn Leu Ile Phe
215 220 225
Lys Tyr Val Gly Thr Met Glu Ala Ser Glu Asp Ala Ala Phe Asn
230 235 240
Lys Ile Thr Arg Ser Leu Gln Asp Leu Gln Gln Lys Asp Ile Gly
245 250 255
Val Lys Pro Glu Phe Ser Phe Asn Ile Pro Arg Ala Lys Arg Glu
260 265 270
Leu Ala Gln Leu Asn Lys Cys Thr Ser Pro Gln Gln Lys Leu Val
275 280 285
Cys Leu Arg Lys Val Val Gln Leu Ile Thr Gln Ser Pro Ser Gln
290 295 300
Arg Val Asn Leu Glu Thr Met Cys Ala Asp Asp Leu Leu Ser Val
305 310 315
Leu Leu Tyr Leu Leu Val Lys Thr Glu Ile Pro Asn Trp Met Ala
320 325 330
Asn Leu Ser Tyr Ile Lys Asn Phe Arg Phe Ser Ser Leu Ala Lys
335 340 345
Asp Glu Leu Gly Tyr Cys Leu Thr Ser Phe Glu Ala Ala Ile Glu
350 355 360
Tyr Ile Arg Gln Gly Ser Leu Ser Ala Lys Pro Pro Glu Ser Glu
365 370 375
Gly Phe Gly Asp Arg Leu Phe Leu Lys Gln Arg Met Ser Leu Leu
380 385 390
Ser Gln Met Thr Ser Ser Pro Thr Asp Cys Leu Phe Lys His Ile
395 400 405
Ala Ser Gly Asn Gln Lys Glu Val Glu Arg Leu Leu Ser Gln Glu
410 415 420
Asp His Asp Lys Asp Thr Val Gln Lys Met Cys His Pro Leu Cys
425 430 435
Phe Cys Asp Asp Cys Glu Lys Leu Val Ser Gly Arg Leu Asn Asp
440 445 450
Pro Ser Val Val Thr Pro Phe Ser Arg Asp Asp Arg Gly His Thr
455 460 465
Pro Leu His Val Ala Ala Val Cys Gly Gln Ala Ser Leu Ile Asp
470 475 480

```
Leu Leu Val Ser Lys Gly Ala Met Val Asn Ala Thr Asp Tyr His
                485                 490                 495
Gly Ala Thr Pro Leu His Leu Ala Cys Gln Lys Gly Tyr Gln Ser
                500                 505                 510
Val Thr Leu Leu Leu Leu His Tyr Lys Ala Ser Ala Glu Val Gln
                515                 520                 525
Asp Asn Asn Gly Asn Thr Pro Leu His Leu Ala Cys Thr Tyr Gly
                530                 535                 540
His Glu Asp Cys Val Lys Ala Leu Val Tyr Tyr Asp Val Glu Ser
                545                 550                 555
Cys Arg Leu Asp Ile Gly Asn Glu Lys Gly Asp Thr Pro Leu His
                560                 565                 570
Ile Ala Ala Arg Trp Gly Tyr Gln Gly Val Ile Glu Thr Leu Leu
                575                 580                 585
Gln Asn Gly Ala Ser Thr Glu Ile Gln Asn Arg Leu Lys Glu Thr
                590                 595                 600
Pro Leu Lys Cys Ala Leu Asn Ser Lys Ile Leu Ser Val Met Glu
                605                 610                 615
Ala Tyr His Leu Ser Phe Glu Arg Arg Gln Lys Ser Ser Glu Ala
                620                 625                 630
Pro Val Gln Ser Pro Gln Arg Ser Val Asp Ser Ile Ser Gln Glu
                635                 640                 645
Ser Ser Thr Ser Ser Phe Ser Ser Met Ser Ala Ser Ser Arg Gln
                650                 655                 660
Glu Glu Thr Lys Lys Asp Tyr Arg Glu Val Glu Lys Leu Leu Arg
                665                 670                 675
Ala Val Ala Asp Gly Asp Leu Glu Met Val Arg Tyr Leu Leu Glu
                680                 685                 690
Trp Thr Glu Glu Asp Leu Glu Asp Ala Glu Asp Thr Val Ser Ala
                695                 700                 705
Ala Asp Pro Glu Phe Cys His Pro Leu Cys Gln Cys Pro Lys Cys
                710                 715                 720
Ala Pro Ala Gln Lys Arg Leu Ala Lys Val Pro Ala Ser Gly Leu
                725                 730                 735
Gly Val Asn Val Thr Ser Gln Asp Gly Ser Ser Pro Leu His Val
                740                 745                 750
Ala Ala Leu His Gly Arg Ala Asp Leu Ile Pro Leu Leu Leu Lys
                755                 760                 765
His Gly Ala Asn Ala Gly Ala Arg Asn Ala Asp Gln Ala Val Pro
                770                 775                 780
Leu His Leu Ala Cys Gln Gln Gly His Phe Gln Val Val Lys Cys
                785                 790                 795
Leu Leu Asp Ser Asn Ala Lys Pro Asn Lys Lys Asp Leu Ser Gly
                800                 805                 810
Asn Thr Pro Leu Ile Tyr Ala Cys Ser Gly Gly His His Glu Leu
                815                 820                 825
Val Ala Leu Leu Leu Gln His Gly Ala Ser Ile Asn Ala Ser Asn
                830                 835                 840
Asn Lys Gly Asn Thr Ala Leu His Glu Ala Val Ile Glu Lys His
                845                 850                 855
Val Phe Val Val Glu Leu Leu Leu Leu His Gly Ala Ser Val Gln
                860                 865                 870
Val Leu Asn Lys Arg Gln Arg Thr Ala Val Asp Cys Ala Glu Gln
                875                 880                 885
Asn Ser Lys Ile Met Glu Leu Leu Gln Val Val Pro Ser Cys Val
                890                 895                 900
```

```
Ala Ser Leu Asp Asp Val Ala Glu Thr Asp Arg Lys Glu Tyr Val
                905                 910                 915
Thr Val Lys Ile Arg Lys Lys Trp Asn Ser Lys Leu Tyr Asp Leu
                920                 925                 930
Pro Asp Glu Pro Phe Thr Arg Gln Phe Tyr Phe Val His Ser Ala
                935                 940                 945
Gly Gln Phe Lys Gly Lys Thr Ser Arg Glu Ile Met Ala Arg Asp
                950                 955                 960
Arg Ser Val Pro Asn Leu Thr Glu Gly Ser Leu His Glu Pro Gly
                965                 970                 975
Arg Gln Ser Val Thr Leu Arg Gln Asn Asn Leu Pro Ala Gln Ser
                980                 985                 990
Gly Ser His Ala Ala Glu Lys Gly Asn Ser Asp Trp Pro Glu Arg
                995                1000                1005
Pro Gly Leu Thr Gln Thr Gly Pro Gly His Arg Arg Met Leu Arg
               1010                1015                1020
Arg His Thr Val Glu Asp Ala Val Val Ser Gln Gly Pro Glu Ala
               1025                1030                1035
Ala Gly Pro Leu Ser Thr Pro Gln Glu Val Ser Ala Ser Arg Ser
               1040                1045                1050


<210> 4
<211> 326
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7131221CD1

<400> 4
Met Asn Phe Thr Val Gly Phe Lys Pro Leu Leu Gly Asp Ala His
  1               5                  10                  15
Ser Met Asp Asn Leu Glu Lys Gln Leu Ile Cys Pro Ile Cys Leu
                 20                  25                  30
Glu Met Phe Ser Lys Pro Val Val Ile Leu Pro Cys Gln His Asn
                 35                  40                  45
Leu Cys Arg Lys Cys Ala Asn Asp Val Phe Gln Ala Ser Asn Pro
                 50                  55                  60
Leu Trp Gln Ser Arg Gly Ser Thr Thr Val Ser Ser Gly Gly Arg
                 65                  70                  75
Phe Arg Cys Pro Ser Cys Arg His Glu Val Val Leu Asp Arg His
                 80                  85                  90
Gly Val Tyr Gly Leu Gln Arg Asn Leu Leu Val Glu Asn Ile Ile
                 95                 100                 105
Asp Ile Tyr Lys Gln Glu Ser Ser Arg Pro Leu His Ser Lys Ala
                110                 115                 120
Glu Gln His Leu Met Cys Glu Glu His Glu Glu Glu Lys Ile Asn
                125                 130                 135
Ile Tyr Cys Leu Ser Cys Glu Val Pro Thr Cys Ser Leu Cys Lys
                140                 145                 150
Val Phe Gly Ala His Lys Asp Cys Glu Val Ala Pro Leu Pro Thr
                155                 160                 165
Ile Tyr Lys Arg Gln Lys Asp Asn Ser Arg Arg Gln Lys Gln Leu
                170                 175                 180
```

```
Leu Asn Gln Arg Phe Glu Ser Leu Cys Ala Val Leu Glu Glu Arg
                185                 190                 195
Lys Gly Glu Leu Leu Gln Ala Leu Ala Arg Glu Gln Glu Glu Lys
                200                 205                 210
Leu Gln Arg Val Arg Gly Leu Ile Arg Gln Tyr Gly Asp His Leu
                215                 220                 225
Glu Ala Ser Ser Lys Leu Val Glu Ser Ala Ile Gln Ser Met Glu
                230                 235                 240
Glu Pro Gln Met Ala Leu Tyr Leu Gln Gln Ala Lys Glu Leu Ile
                245                 250                 255
Asn Lys Val Gly Ala Met Ser Lys Val Glu Leu Ala Gly Arg Pro
                260                 265                 270
Glu Pro Gly Tyr Glu Ser Met Glu Gln Phe Thr Val Arg Val Glu
                275                 280                 285
His Val Ala Glu Met Leu Arg Thr Ile Asp Phe Gln Pro Gly Ala
                290                 295                 300
Ser Gly Glu Glu Glu Glu Val Ala Pro Asp Gly Glu Glu Gly Ser
                305                 310                 315
Ala Gly Pro Glu Glu Glu Arg Pro Asp Gly Pro
                320                 325

<210> 5
<211> 505
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480551CD1

<400> 5
Met Leu Ser Phe Phe Arg Arg Thr Leu Gly Arg Arg Ser Met Arg
  1               5                  10                  15
Lys His Ala Glu Lys Glu Arg Leu Arg Glu Ala Gln Arg Ala Ala
                 20                  25                  30
Thr His Ile Pro Ala Ala Gly Asp Ser Lys Ser Ile Ile Thr Cys
                 35                  40                  45
Arg Val Ser Leu Leu Asp Gly Thr Asp Val Ser Val Asp Leu Pro
                 50                  55                  60
Lys Lys Ala Lys Gly Gln Glu Leu Phe Asp Gln Ile Met Tyr His
                 65                  70                  75
Leu Asp Leu Ile Glu Ser Asp Tyr Phe Gly Leu Arg Phe Met Asp
                 80                  85                  90
Ser Ala Gln Val Ala His Trp Leu Asp Gly Thr Lys Ser Ile Lys
                 95                 100                 105
Lys Gln Val Lys Ile Gly Ser Pro Tyr Cys Leu His Leu Arg Val
                110                 115                 120
Lys Phe Tyr Ser Ser Glu Pro Asn Asn Leu Arg Glu Glu Leu Thr
                125                 130                 135
Arg Tyr Leu Phe Val Leu Gln Leu Lys Gln Asp Ile Leu Ser Gly
                140                 145                 150
Lys Leu Asp Cys Pro Phe Asp Thr Ala Val Gln Leu Ala Ala Tyr
                155                 160                 165
Asn Leu Gln Ala Glu Leu Gly Asp Tyr Asp Leu Ala Glu His Ser
                170                 175                 180
Pro Glu Leu Val Ser Glu Phe Arg Phe Val Pro Ile Gln Thr Glu
```

                185   190   195
Glu Met Glu Leu Ala Ile Phe Glu Lys Trp Lys Glu Tyr Arg Gly
                200   205   210
Gln Thr Pro Ala Gln Ala Glu Thr Asn Tyr Leu Asn Lys Ala Lys
                215   220   225
Trp Leu Glu Met Tyr Gly Val Asp Met His Val Val Lys Ala Arg
                230   235   240
Asp Gly Asn Asp Tyr Ser Leu Gly Leu Thr Pro Thr Gly Val Leu
                245   250   255
Val Phe Glu Gly Asp Thr Lys Ile Gly Leu Phe Phe Trp Pro Lys
                260   265   270
Ile Thr Arg Leu Asp Phe Lys Lys Asn Lys Leu Thr Leu Val Val
                275   280   285
Val Glu Asp Asp Asp Gln Gly Lys Glu Gln Glu His Thr Phe Val
                290   295   300
Phe Arg Leu Asp His Pro Lys Ala Cys Lys His Leu Trp Lys Cys
                305   310   315
Ala Val Glu His His Ala Phe Phe Arg Leu Arg Gly Pro Val Gln
                320   325   330
Lys Ser Ser His Arg Ser Gly Phe Ile Arg Leu Gly Ser Arg Phe
                335   340   345
Arg Tyr Ser Gly Lys Thr Glu Tyr Gln Thr Thr Lys Thr Asn Lys
                350   355   360
Ala Arg Arg Ser Thr Ser Phe Glu Arg Arg Pro Ser Lys Arg Tyr
                365   370   375
Ser Arg Arg Thr Leu Gln Met Lys Ala Cys Ala Thr Lys Pro Glu
                380   385   390
Glu Leu Ser Val His Asn Asn Val Ser Thr Gln Ser Asn Gly Ser
                395   400   405
Gln Gln Ala Trp Gly Met Arg Ser Ala Leu Pro Val Ser Pro Ser
                410   415   420
Ile Ser Ser Ala Pro Val Pro Val Glu Ile Glu Asn Leu Pro Gln
                425   430   435
Ser Pro Gly Thr Asp Gln His Asp Arg Lys Trp Leu Ser Ala Ala
                440   445   450
Ser Asp Cys Cys Gln Arg Gly Gly Asn Gln Trp Asn Thr Arg Ala
                455   460   465
Leu Pro Pro Pro Gln Thr Ala His Arg Asn Tyr Thr Asp Phe Val
                470   475   480
His Glu His Asn Val Lys Asn Ala Gly Ile Arg His Asp Val His
                485   490   495
Phe Pro Gly His Thr Ala Met Thr Glu Ile
                500   505

<210> 6
<211> 367
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3315870CD1

<400> 6
Met Ala Val Leu Lys Leu Thr Asp Gln Pro Pro Leu Val Gln Ala
  1   5   10   15

```
Ile Phe Ser Gly Asp Pro Glu Glu Ile Arg Met Leu Ile His Lys
                20                  25                  30
Thr Glu Asp Val Asn Thr Leu Asp Ser Glu Lys Arg Thr Pro Leu
                35                  40                  45
His Val Ala Ala Phe Leu Gly Asp Ala Glu Ile Ile Glu Leu Leu
                50                  55                  60
Ile Leu Ser Gly Ala Arg Val Asn Ala Lys Asp Asn Met Trp Leu
                65                  70                  75
Thr Pro Leu His Arg Ala Val Ala Ser Arg Ser Glu Glu Ala Val
                80                  85                  90
Gln Val Leu Ile Lys His Ser Ala Asp Val Asn Ala Arg Asp Lys
                95                  100                 105
Asn Trp Gln Thr Pro Leu His Val Ala Ala Ala Asn Lys Ala Val
                110                 115                 120
Lys Cys Ala Glu Val Ile Ile Pro Leu Leu Ser Ser Val Asn Val
                125                 130                 135
Ser Asp Arg Gly Gly Arg Thr Ala Leu His His Ala Ala Leu Asn
                140                 145                 150
Gly His Val Glu Met Val Asn Leu Leu Leu Ala Lys Gly Ala Asn
                155                 160                 165
Ile Asn Ala Phe Asp Lys Lys Asp Arg Arg Ala Leu His Trp Ala
                170                 175                 180
Ala Tyr Met Gly His Leu Asp Val Val Ala Leu Leu Ile Asn His
                185                 190                 195
Gly Ala Glu Val Thr Cys Lys Asp Lys Lys Gly Tyr Thr Pro Leu
                200                 205                 210
His Ala Ala Ala Ser Asn Gly Gln Ile Asn Val Val Lys His Leu
                215                 220                 225
Leu Asn Leu Gly Val Glu Ile Asp Glu Ile Asn Val Tyr Gly Asn
                230                 235                 240
Thr Ala Leu His Ile Ala Cys Tyr Asn Gly Gln Asp Ala Val Val
                245                 250                 255
Asn Glu Leu Ile Asp Tyr Gly Ala Asn Val Asn Gln Pro Asn Asn
                260                 265                 270
Asn Gly Phe Thr Pro Leu His Phe Ala Ala Ala Ser Thr His Gly
                275                 280                 285
Ala Leu Cys Leu Glu Leu Leu Val Asn Asn Gly Ala Asp Val Asn
                290                 295                 300
Ile Gln Ser Lys Asp Gly Lys Ser Pro Leu His Met Thr Ala Val
                305                 310                 315
His Gly Arg Phe Thr Arg Ser Gln Thr Leu Ile Gln Asn Gly Gly
                320                 325                 330
Glu Ile Asp Cys Val Asp Lys Asp Gly Asn Thr Pro Leu His Val
                335                 340                 345
Ala Ala Arg Tyr Gly His Glu Leu Leu Ile Asn Thr Leu Ile Thr
                350                 355                 360
Ser Gly Ala Asp Thr Ala Lys
                365
```

<210> 7
<211> 435
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7484690CD1

<400> 7

Met Arg Glu Ile Val Leu Thr Gln Thr Gly Gln Cys Gly Asn Gln
1 5 10 15
Ile Gly Ala Lys Gln Phe Trp Glu Val Ile Ser Asp Glu His Ala
20 25 30
Ile Asp Ser Ala Gly Thr Tyr His Gly Asp Ser His Leu Pro Leu
35 40 45
Glu Arg Val Asn Val His His His Glu Ala Ser Gly Gly Arg Tyr
50 55 60
Val Pro Arg Ala Val Leu Val Asp Leu Glu Pro Gly Thr Met Asp
65 70 75
Ser Val Arg Ser Gly Pro Phe Gly Gln Val Phe Arg Pro Asp Asn
80 85 90
Phe Ile Ser Arg Gln Cys Gly Ala Gly Asn Asn Trp Ala Lys Gly
95 100 105
Arg Tyr Thr Glu Gly Ala Glu Leu Thr Glu Ser Val Met Asp Val
110 115 120
Val Arg Lys Glu Ala Glu Ser Cys Asp Cys Leu Gln Gly Phe Gln
125 130 135
Leu Thr His Ser Leu Gly Gly Gly Thr Gly Ser Gly Met Gly Thr
140 145 150
Leu Leu Leu Ser Lys Ile Arg Glu Glu Tyr Pro Asp Arg Ile Ile
155 160 165
Asn Thr Phe Ser Ile Leu Pro Ser Pro Lys Val Ser Asp Thr Val
170 175 180
Val Glu Pro Tyr Asn Val Thr Leu Ser Val His Gln Leu Ile Glu
185 190 195
Asn Ala Asp Glu Thr Phe Cys Ile Asp Asn Glu Ala Leu Tyr Asp
200 205 210
Ile Cys Ser Arg Thr Leu Lys Leu Pro Thr Pro Thr Tyr Gly Asp
215 220 225
Leu Asn His Leu Val Ser Ala Thr Met Ser Gly Val Thr Thr Cys
230 235 240
Leu Arg Phe Pro Gly Gln Leu Asn Ala Asp Leu Arg Lys Leu Ala
245 250 255
Val Asn Met Val Pro Phe Pro Arg Leu His Phe Phe Met Pro Gly
260 265 270
Phe Ala Pro Leu Thr Ser Arg Gly Ser Gln Gln Tyr Arg Ala Leu
275 280 285
Thr Val Ala Glu Leu Thr Gln Gln Met Phe Asp Ala Lys Asn Met
290 295 300
Met Ala Ala Arg Asp Pro Cys His Gly Arg Tyr Leu Thr Val Ala
305 310 315
Ala Ile Phe Arg Gly Arg Met Pro Met Arg Glu Val Asp Glu Gln
320 325 330
Met Phe Asn Ile Gln Asp Lys Asn Ser Ser Tyr Phe Ala Asp Trp
335 340 345
Phe Pro Asp Asn Val Lys Thr Ala Val Cys Asp Ile Pro Pro Arg
350 355 360
Gly Leu Lys Met Ser Ala Thr Phe Ile Gly Asn Asn Thr Ala Val
365 370 375
Gln Glu Leu Lys Arg Val Ser Glu Gln Phe Thr Ala Thr Phe Arg
380 385 390
Arg Lys Ala Phe Leu His Trp Tyr Thr Gly Glu Gly Met Asp Glu

```
                395                 400                 405
Met Glu Phe Thr Glu Ala Glu Ser Asn Met Asn Asp Leu Val Ser
                410                 415                 420
Glu Tyr Gln Gln Tyr Gln Asp Ala Thr Ala Glu Gly Gly Gly Val
                425                 430                 435
```

<210> 8
<211> 198
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7612559CD1

<400> 8

```
Met Gly Gly Arg Lys Arg Glu Arg Lys Ala Ala Val Glu Glu Asp
  1               5                  10                  15
Thr Ser Leu Ser Glu Ser Glu Gly Pro Arg Gln Pro Asp Gly Asp
                 20                  25                  30
Glu Glu Glu Ser Thr Ala Leu Ser Ile Asn Glu Glu Met Gln Arg
                 35                  40                  45
Met Leu Asn Gln Leu Arg Glu Tyr Asp Phe Glu Asp Asp Cys Asp
                 50                  55                  60
Ser Leu Thr Trp Glu Glu Thr Glu Glu Thr Leu Leu Leu Trp Glu
                 65                  70                  75
Asp Phe Ser Gly Tyr Ala Met Ala Ala Ala Glu Ala Gln Gly Glu
                 80                  85                  90
Gln Gln Glu Asp Ser Leu Glu Lys Val Ile Lys Asp Thr Glu Ser
                 95                 100                 105
Leu Phe Lys Thr Arg Glu Lys Glu Tyr Gln Glu Thr Ile Asp Gln
                110                 115                 120
Ile Glu Leu Glu Leu Ala Thr Ala Lys Asn Asp Met Asn Arg His
                125                 130                 135
Leu His Glu Tyr Met Glu Met Cys Ser Met Lys Arg Gly Leu Asp
                140                 145                 150
Val Gln Met Glu Thr Cys Arg Arg Leu Ile Thr Gln Ser Gly Asp
                155                 160                 165
Arg Lys Ser Pro Ala Phe Thr Ala Val Pro Leu Ser Asp Arg Arg
                170                 175                 180
Arg Arg Gln Ala Arg Leu Arg Thr Pro Ile Ala Met Ser His Leu
                185                 190                 195
Thr Ala Pro
```

<210> 9
<211> 139
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4940751CD1

<400> 9

Met Ala Asn Ala Arg Ser Gly Val Ala Val Asn Asp Glu Cys Met
1   5   10   15
Leu Lys Phe Gly Glu Leu Gln Ser Lys Arg Leu His Arg Phe Leu
                20   25   30
Thr Phe Lys Met Asp Asp Lys Phe Lys Glu Ile Val Val Asp Gln
                35   40   45
Val Gly Asp Arg Ala Thr Ser Tyr Glu Asp Phe Thr Asn Ser Leu
                50   55   60
Pro Glu Asn Asp Cys Arg Tyr Ala Ile Tyr Asp Phe Asp Phe Val
                65   70   75
Thr Ala Glu Asp Val Gln Lys Ser Arg Ile Phe Tyr Ile Leu Trp
                80   85   90
Ser Pro Ser Ser Ala Lys Val Lys Ser Lys Met Leu Tyr Ala Ser
                95   100   105
Ser Asn Gln Lys Phe Lys Ser Gly Leu Asn Gly Ile Gln Val Glu
                110   115   120
Leu Gln Ala Thr Asp Ala Ser Glu Ile Ser Leu Asp Glu Ile Lys
                125   130   135
Asp Arg Ala Arg

<210> 10
<211> 736
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7946761CD1

<400> 10
Met Thr Trp Gly Thr Pro Asp Phe Leu Asn Arg Ser Ser Thr His
1   5   10   15
Ser Ser Arg Val Pro Ser Arg Phe Pro Phe Leu Asn Glu Ile Val
                20   25   30
Ala His Pro Val Ala Ser Ser His Pro Gly Ser Tyr Arg Arg Ser
                35   40   45
Gln Thr Leu Leu Glu Arg Leu Arg Val Ser Arg Ala Pro Glu Asp
                50   55   60
Thr Lys Ala Leu Glu Pro Arg Cys Gly Pro Pro Cys Gly Ala Gly
                65   70   75
Gln Pro Gly Trp Glu Pro Cys Ser Ala Leu Glu Arg Gly Pro Pro
                80   85   90
Ser Arg Gly Glu Glu Arg Arg Met Pro Thr Ser Pro Pro Ala Gly
                95   100   105
Ser Arg Lys Ser Thr Asp Gln Ala Val Arg Phe Gly Pro Ser Gln
                110   115   120
Gly Met Cys Ser Glu Ala Arg Leu Ala Arg Arg Leu Arg Asp Ala
                125   130   135
Leu Arg Glu Glu Glu Pro Trp Ala Val Glu Glu Leu Leu Arg Cys
                140   145   150
Gly Ala Asp Pro Asn Leu Val Leu Glu Asp Gly Ala Ala Ala Val
                155   160   165
His Leu Ala Ala Gly Ala Arg His Pro Arg Gly Leu Arg Cys Leu
                170   175   180
Gly Ala Leu Leu Arg Gln Gly Gly Asp Pro Asn Ala Arg Ser Val

```
                185                 190                 195
Glu Ala Leu Thr Pro Leu His Val Ala Ala Ala Trp Gly Cys Arg
                200                 205                 210
Arg Gly Leu Glu Leu Leu Leu Ser Gln Gly Ala Asp Pro Ala Leu
                215                 220                 225
Arg Asp Gln Asp Gly Leu Arg Pro Leu Asp Leu Ala Leu Gln Gln
                230                 235                 240
Gly His Leu Glu Cys Ala Arg Val Leu Gln Asp Leu Asp Thr Arg
                245                 250                 255
Thr Arg Thr Arg Thr Arg Ile Gly Ala Glu Thr Gln Glu Pro Glu
                260                 265                 270
Pro Ala Pro Gly Thr Pro Gly Leu Ser Gly Pro Thr Asp Glu Thr
                275                 280                 285
Leu Asp Ser Ile Ala Leu Gln Lys Gln Pro Cys Arg Gly Asp Asn
                290                 295                 300
Arg Asp Ile Gly Leu Glu Ala Asp Pro Gly Pro Pro Ser Leu Pro
                305                 310                 315
Val Pro Leu Glu Thr Val Asp Lys His Gly Ser Ser Ala Ser Pro
                320                 325                 330
Pro Gly His Trp Asp Tyr Ser Ser Asp Ala Ser Phe Val Thr Ala
                335                 340                 345
Val Glu Val Ser Gly Ala Glu Asp Pro Ala Ser Asp Thr Pro Pro
                350                 355                 360
Trp Ala Gly Ser Leu Pro Pro Thr Arg Gln Gly Leu Leu His Val
                365                 370                 375
Val His Ala Asn Gln Arg Val Pro Arg Ser Gln Gly Thr Glu Ala
                380                 385                 390
Glu Leu Asn Ala Arg Leu Gln Ala Leu Thr Leu Thr Pro Pro Asn
                395                 400                 405
Ala Ala Gly Phe Gln Ser Ser Pro Ser Ser Met Pro Leu Leu Asp
                410                 415                 420
Arg Ser Pro Ala His Ser Pro Pro Arg Thr Pro Thr Pro Gly Ala
                425                 430                 435
Ser Asp Cys His Cys Leu Trp Glu His Gln Thr Ser Ile Asp Ser
                440                 445                 450
Asp Met Ala Thr Leu Trp Leu Thr Glu Asp Glu Ala Ser Ser Thr
                455                 460                 465
Gly Gly Arg Glu Pro Val Gly Pro Cys Arg His Leu Pro Val Ser
                470                 475                 480
Thr Val Ser Asp Leu Glu Leu Leu Lys Gly Leu Arg Ala Leu Gly
                485                 490                 495
Glu Asn Pro His Pro Ile Thr Pro Phe Thr Arg Gln Leu Tyr His
                500                 505                 510
Gln Gln Leu Glu Glu Ala Gln Ile Ala Pro Gly Pro Glu Phe Ser
                515                 520                 525
Gly His Ser Leu Glu Leu Ala Ala Ala Leu Arg Thr Gly Cys Ile
                530                 535                 540
Pro Asp Val Gln Ala Asp Glu Asp Ala Leu Ala Gln Gln Phe Glu
                545                 550                 555
Arg Pro Asp Pro Ala Arg Arg Trp Arg Glu Gly Val Val Lys Ser
                560                 565                 570
Ser Phe Thr Tyr Leu Leu Leu Asp Pro Arg Glu Thr Gln Asp Leu
                575                 580                 585
Pro Ala Arg Ala Phe Ser Leu Thr Pro Ala Glu Arg Leu Gln Thr
                590                 595                 600
Phe Ile Arg Ala Ile Phe Tyr Val Gly Lys Gly Thr Arg Ala Arg
```

```
                605                 610                 615
Pro Tyr Val His Leu Trp Glu Ala Leu Gly His His Gly Arg Ser
                620                 625                 630
Arg Lys Gln Pro His Gln Ala Cys Pro Lys Val Arg Gln Ile Leu
                635                 640                 645
Asp Ile Trp Ala Ser Gly Cys Gly Val Val Ser Leu His Cys Phe
                650                 655                 660
Gln His Val Val Ala Val Glu Ala Tyr Thr Arg Glu Ala Cys Ile
                665                 670                 675
Val Glu Ala Leu Gly Ile Gln Thr Leu Thr Asn Gln Lys Gln Gly
                680                 685                 690
His Cys Tyr Gly Val Val Ala Gly Trp Pro Pro Ala Arg Arg Arg
                695                 700                 705
Arg Leu Gly Val His Leu Leu His Arg Ala Leu Leu Val Phe Leu
                710                 715                 720
Ala Glu Gly Glu Arg Gln Leu His Pro Gln Asp Ile Gln Ala Arg
                725                 730                 735
Gly


<210> 11
<211> 529
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3288747CD1

<400> 11
Met Ser Arg Gln Phe Thr Tyr Lys Ser Gly Ala Ala Ala Lys Gly
  1               5                  10                  15
Gly Phe Ser Gly Cys Ser Ala Val Leu Ser Gly Gly Ser Ser Ser
                 20                  25                  30
Ser Tyr Arg Ala Gly Gly Lys Gly Leu Ser Gly Gly Phe Ser Ser
                 35                  40                  45
Arg Ser Leu Tyr Ser Leu Gly Gly Ala Arg Ser Ile Ser Phe Asn
                 50                  55                  60
Val Ala Ser Gly Ser Gly Trp Ala Gly Gly Tyr Gly Phe Gly Arg
                 65                  70                  75
Gly Arg Ala Ser Gly Phe Ala Gly Ser Met Phe Gly Ser Val Ala
                 80                  85                  90
Leu Gly Ser Val Cys Pro Ser Leu Cys Pro Pro Gly Gly Ile His
                 95                 100                 105
Gln Val Thr Ile Asn Lys Ser Leu Leu Ala Pro Leu Asn Val Glu
                110                 115                 120
Leu Asp Pro Glu Ile Gln Lys Val Arg Ala Gln Glu Arg Glu Gln
                125                 130                 135
Ile Lys Val Leu Asn Asn Lys Phe Ala Ser Phe Ile Asp Lys Val
                140                 145                 150
Arg Phe Leu Glu Gln Gln Asn Gln Val Leu Glu Thr Lys Trp Glu
                155                 160                 165
Leu Leu Gln Gln Leu Asp Leu Asn Asn Cys Lys Asn Asn Leu Glu
                170                 175                 180
Pro Ile Leu Glu Gly Tyr Ile Ser Asn Leu Arg Lys Gln Leu Glu
                185                 190                 195
```

```
Thr Leu Ser Gly Asp Arg Val Arg Leu Asp Ser Glu Leu Arg Ser
                200                 205                 210
Val Arg Glu Val Val Glu Asp Tyr Lys Lys Arg Tyr Glu Glu Glu
                215                 220                 225
Ile Asn Lys Arg Thr Thr Ala Glu Asn Glu Phe Val Val Leu Lys
                230                 235                 240
Lys Asp Val Asp Ala Ala Tyr Thr Ser Lys Val Glu Leu Gln Ala
                245                 250                 255
Lys Val Asp Ala Leu Asp Gly Glu Ile Lys Phe Phe Lys Cys Leu
                260                 265                 270
Tyr Glu Gly Glu Thr Ala Gln Ile Gln Ser His Ile Ser Asp Thr
                275                 280                 285
Ser Ile Ile Leu Ser Met Asp Asn Asn Arg Asn Leu Asp Leu Asp
                290                 295                 300
Ser Ile Ile Ala Glu Val Arg Ala Gln Tyr Glu Glu Ile Ala Arg
                305                 310                 315
Lys Ser Lys Ala Glu Ala Glu Ala Leu Tyr Gln Thr Lys Phe Gln
                320                 325                 330
Glu Leu Gln Leu Ala Ala Gly Arg His Gly Asp Asp Leu Lys His
                335                 340                 345
Thr Lys Asn Glu Ile Ser Glu Leu Thr Arg Leu Ile Gln Arg Leu
                350                 355                 360
Arg Ser Glu Ile Glu Ser Val Lys Lys Gln Cys Ala Asn Leu Glu
                365                 370                 375
Thr Ala Ile Ala Asp Ala Glu Gln Arg Gly Asp Cys Ala Leu Lys
                380                 385                 390
Asp Ala Arg Ala Lys Leu Asp Glu Leu Glu Gly Ala Leu Gln Gln
                395                 400                 405
Ala Lys Glu Glu Leu Ala Arg Met Leu Arg Glu Tyr Gln Glu Leu
                410                 415                 420
Leu Ser Val Lys Leu Ser Leu Asp Ile Glu Ile Ala Thr Tyr Arg
                425                 430                 435
Lys Leu Leu Glu Gly Glu Glu Cys Arg Met Ser Gly Glu Tyr Thr
                440                 445                 450
Asn Ser Val Ser Ile Ser Val Ile Asn Ser Ser Met Ala Gly Met
                455                 460                 465
Ala Gly Thr Gly Ala Gly Phe Gly Phe Ser Asn Ala Gly Thr Tyr
                470                 475                 480
Gly Tyr Trp Pro Ser Ser Val Ser Gly Gly Tyr Ser Met Leu Pro
                485                 490                 495
Gly Gly Cys Val Thr Gly Ser Gly Asn Cys Ser Pro Pro Val Val
                500                 505                 510
Ser Asn Val Thr Ser Thr Ser Gly Ser Ser Gly Ser Ser Arg Gly
                515                 520                 525
Val Phe Gly Gly
```

<210> 12
<211> 1367
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8200016CD1

<400> 12

Met Ser His Tyr His Phe Ile Lys Cys Cys Cys Phe Gln Leu Cys
1 5 10 15
Asn Val Phe Arg Ser His Glu Met Glu Ile Asp Gln Cys Leu Leu
20 25 30
Glu Ser Leu Pro Leu Gly Gln Arg Gln Arg Leu Val Lys Arg Met
35 40 45
Arg Cys Glu Gln Ile Lys Ala Tyr Tyr Glu Arg Glu Lys Ala Phe
50 55 60
Gln Lys Gln Glu Gly Phe Leu Lys Arg Leu Lys His Ala Lys Asn
65 70 75
Pro Lys Val His Phe Asn Leu Thr Asp Met Leu Gln Asp Ala Ile
80 85 90
Ile His His Asn Asp Lys Glu Val Leu Arg Leu Leu Lys Glu Gly
95 100 105
Ala Asp Pro His Thr Leu Val Ser Ser Gly Gly Ser Leu Leu His
110 115 120
Leu Cys Ala Arg Tyr Asp Asn Ala Phe Ile Ala Glu Ile Leu Ile
125 130 135
Asp Arg Gly Val Asn Val Asn His Gln Asp Glu Asp Phe Trp Thr
140 145 150
Pro Met His Ile Ala Cys Ala Cys Asp Asn Pro Asp Ile Val Leu
155 160 165
Leu Leu Val Leu Ala Gly Ala Asn Val Leu Leu Gln Asp Val Asn
170 175 180
Gly Asn Ile Pro Leu Asp Tyr Ala Val Glu Gly Thr Glu Ser Ser
185 190 195
Ser Ile Leu Leu Thr Tyr Leu Asp Glu Asn Gly Val Asp Leu Thr
200 205 210
Ser Leu Arg Gln Met Lys Leu Gln Arg Pro Met Ser Met Leu Thr
215 220 225
Asp Val Lys His Phe Leu Ser Ser Gly Gly Asn Val Asn Glu Lys
230 235 240
Asn Asp Glu Gly Val Thr Leu Leu His Met Ala Cys Ala Ser Gly
245 250 255
Tyr Lys Glu Val Val Ser Leu Ile Leu Glu His Gly Gly Asp Leu
260 265 270
Asn Ile Val Asp Asp Gln Tyr Trp Thr Pro Leu His Leu Ala Ala
275 280 285
Lys Tyr Gly Gln Thr Asn Leu Val Lys Leu Leu Leu Met His Gln
290 295 300
Ala Asn Pro His Leu Val Asn Cys Asn Glu Glu Lys Ala Ser Asp
305 310 315
Ile Ala Ala Ser Glu Phe Ile Glu Glu Met Leu Leu Lys Ala Glu
320 325 330
Ile Ala Trp Glu Glu Lys Met Lys Glu Pro Leu Ser Ala Ser Thr
335 340 345
Leu Ala Gln Glu Glu Pro Tyr Glu Glu Ile Ile His Asp Leu Pro
350 355 360
Val Leu Ser Ser Lys Leu Ser Pro Leu Val Leu Pro Ile Ala Lys
365 370 375
Gln Asp Ser Leu Leu Glu Lys Asp Ile Met Phe Lys Asp Ala Thr
380 385 390
Lys Gly Leu Cys Lys Gln Gln Ser Gln Asp Ser Ile Pro Glu Asn
395 400 405
Pro Met Met Ser Gly Ser Thr Lys Pro Glu Gln Val Lys Leu Met

410 415 420
Pro Pro Ala Pro Asn Asp Asp Leu Ala Thr Leu Ser Glu Leu Asn
425 430 435
Asp Gly Ser Leu Leu Tyr Glu Ile Gln Lys Arg Phe Gly Asn Asn
440 445 450
Gln Ile Tyr Thr Phe Ile Gly Asp Ile Leu Leu Leu Val Asn Pro
455 460 465
Tyr Lys Glu Leu Pro Ile Tyr Ser Ser Met Val Ser Gln Leu Tyr
470 475 480
Phe Ser Ser Ser Gly Lys Leu Cys Ser Ser Leu Pro Pro His Leu
485 490 495
Phe Ser Cys Val Glu Arg Ala Phe His Gln Leu Phe Arg Glu Gln
500 505 510
Arg Pro Gln Cys Phe Ile Leu Ser Gly Glu Arg Gly Ser Gly Lys
515 520 525
Ser Glu Ala Ser Lys Gln Ile Ile Arg His Leu Thr Cys Arg Ala
530 535 540
Gly Ala Ser Arg Ala Thr Leu Asp Ser Arg Phe Lys His Val Val
545 550 555
Cys Ile Leu Glu Ala Phe Gly His Ala Lys Thr Thr Leu Asn Asp
560 565 570
Leu Ser Ser Cys Phe Ile Lys Tyr Phe Glu Leu Gln Phe Cys Glu
575 580 585
Arg Lys Gln Gln Leu Thr Gly Ala Arg Ile Tyr Thr Tyr Leu Leu
590 595 600
Glu Lys Ser Arg Leu Val Ser Gln Pro Leu Gly Gln Ser Asn Phe
605 610 615
Leu Ile Phe Tyr Leu Leu Met Asp Gly Leu Ser Ala Glu Glu Lys
620 625 630
Tyr Gly Leu His Leu Asn Asn Leu Cys Ala His Arg Tyr Leu Asn
635 640 645
Gln Thr Ile Gln Asp Asp Ala Ser Thr Gly Glu Arg Ser Leu Asn
650 655 660
Arg Glu Lys Leu Ala Val Leu Lys Arg Ala Leu Asn Val Val Gly
665 670 675
Phe Ser Ser Leu Glu Val Glu Asn Leu Phe Val Ile Leu Ala Ala
680 685 690
Ile Leu His Leu Gly Asp Ile Arg Phe Thr Ala Leu Asn Glu Gly
695 700 705
Asn Ser Ala Phe Val Ser Asp Leu Gln Leu Leu Glu Gln Val Ala
710 715 720
Gly Met Leu Gln Val Ser Thr Asp Glu Leu Ala Ser Ala Leu Thr
725 730 735
Thr Asp Ile Gln Tyr Phe Lys Gly Asp Met Ile Ile Arg Arg His
740 745 750
Thr Ile Gln Ile Ala Glu Phe Phe Arg Asp Leu Leu Ala Lys Ser
755 760 765
Leu Tyr Ser Arg Leu Phe Ser Phe Leu Val Asn Thr Met Asn Ser
770 775 780
Cys Leu His Ser Gln Asp Glu Gln Lys Ser Met Gln Thr Leu Asp
785 790 795
Ile Gly Ile Leu Asp Ile Phe Gly Phe Glu Glu Phe Gln Lys Asn
800 805 810
Glu Phe Glu Gln Leu Cys Val Asn Met Thr Asn Glu Lys Met His
815 820 825
His Tyr Ile Asn Glu Val Leu Phe Leu His Glu Gln Val Glu Cys

830 835 840
Val Gln Glu Gly Val Thr Met Glu Thr Ala Tyr Ser Ala Gly Asn
845 850 855
Gln Asn Gly Val Leu Asp Phe Phe Phe Gln Lys Pro Ser Gly Phe
860 865 870
Leu Thr Leu Leu Asp Glu Glu Ser Gln Met Ile Trp Ser Val Glu
875 880 885
Ser Asn Phe Pro Lys Lys Leu Gln Ser Leu Leu Glu Ser Ser Asn
890 895 900
Thr Asn Ala Val Tyr Ser Pro Met Lys Asp Gly Asn Gly Asn Val
905 910 915
Ala Leu Lys Asp His Gly Thr Ala Phe Thr Ile Met His Tyr Ala
920 925 930
Gly Arg Val Met Tyr Asp Val Val Gly Ala Ile Glu Lys Asn Lys
935 940 945
Asp Ser Leu Ser Gln Asn Leu Leu Phe Val Met Lys Thr Ser Glu
950 955 960
Asn Val Val Ile Asn His Leu Phe Gln Ser Lys Leu Ser Gln Thr
965 970 975
Gly Ser Leu Val Ser Ala Tyr Pro Ser Phe Lys Phe Arg Gly His
980 985 990
Lys Ser Ala Leu Leu Ser Lys Lys Met Thr Ala Ser Ser Ile Ile
995 1000 1005
Gly Glu Asn Lys Asn Tyr Leu Glu Leu Ser Lys Leu Leu Lys Lys
1010 1015 1020
Lys Gly Thr Ser Thr Phe Leu Gln Arg Leu Glu Arg Gly Asp Pro
1025 1030 1035
Val Thr Ile Ala Ser Gln Leu Arg Lys Ser Leu Met Asp Ile Ile
1040 1045 1050
Gly Lys Leu Gln Lys Cys Thr Pro His Phe Ile His Cys Ile Arg
1055 1060 1065
Pro Asn Asn Ser Lys Leu Pro Asp Thr Phe Asp Asn Phe Tyr Val
1070 1075 1080
Ser Ala Gln Leu Gln Tyr Ile Gly Val Leu Glu Met Val Lys Ile
1085 1090 1095
Phe Arg Tyr Gly Tyr Pro Val Arg Leu Ser Phe Ser Asp Phe Leu
1100 1105 1110
Ser Arg Tyr Lys Pro Leu Ala Asp Thr Phe Leu Arg Glu Lys Lys
1115 1120 1125
Glu Gln Ser Ala Ala Glu Arg Cys Arg Leu Val Leu Gln Gln Cys
1130 1135 1140
Lys Leu Gln Gly Trp Gln Met Gly Val Arg Lys Val Phe Leu Lys
1145 1150 1155
Tyr Trp His Ala Asp Gln Leu Asn Asp Leu Cys Leu Gln Leu Gln
1160 1165 1170
Arg Lys Ile Ile Thr Cys Gln Lys Val Ile Arg Gly Phe Leu Ala
1175 1180 1185
Arg Gln His Leu Leu Gln Arg Met Ser Ile Arg Gln Gln Glu Val
1190 1195 1200
Thr Ser Ile Asn Ser Phe Leu Gln Asn Thr Glu Asp Met Gly Leu
1205 1210 1215
Lys Thr Tyr Asp Ala Leu Val Ile Gln Asn Ala Ser Asp Ile Ala
1220 1225 1230
Arg Glu Asn Asp Arg Leu Arg Ser Glu Met Asn Ala Pro Tyr His
1235 1240 1245
Lys Glu Lys Leu Glu Val Arg Asn Met Gln Glu Glu Gly Ser Lys

```
                1250                1255                1260
Arg Thr Asp Asp Lys Ser Gly Pro Arg His Phe His Pro Ser Ser
                1265                1270                1275
Met Ser Val Cys Ala Ala Val Asp Gly Leu Gly Gln Cys Leu Val
                1280                1285                1290
Gly Pro Ser Ile Trp Ser Pro Ser Leu His Ser Val Phe Ser Met
                1295                1300                1305
Asp Asp Ser Ser Ser Leu Pro Ser Pro Arg Lys Gln Pro Pro Pro
                1310                1315                1320
Lys Pro Lys Arg Asp Pro Asn Thr Arg Leu Ser Ala Ser Tyr Glu
                1325                1330                1335
Ala Val Ser Ala Cys Leu Ser Ala Ala Arg Glu Ala Ala Asn Glu
                1340                1345                1350
Gly Gln Pro Trp Gly Gly Thr Gln Pro Arg Val Pro Gly Ser Arg
                1355                1360                1365
Met Leu


<210> 13
<211> 929
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3291962CD1

<400> 13
Met Ala Glu Val Glu Ala Val Gln Leu Lys Glu Glu Gly Asn Arg
  1               5                  10                  15
His Phe Gln Leu Gln Asp Tyr Lys Ala Ala Thr Asn Ser Tyr Ser
                 20                  25                  30
Gln Ala Leu Lys Leu Thr Lys Asp Lys Ala Leu Leu Ala Thr Leu
                 35                  40                  45
Tyr Arg Asn Arg Ala Ala Cys Gly Leu Lys Thr Glu Ser Tyr Val
                 50                  55                  60
Gln Ala Ala Ser Asp Ala Ser Arg Ala Ile Asp Ile Asn Ser Ser
                 65                  70                  75
Asp Ile Lys Ala Leu Tyr Arg Arg Cys Gln Ala Leu Glu His Leu
                 80                  85                  90
Gly Lys Leu Asp Gln Ala Phe Lys Asp Val Gln Arg Cys Ala Thr
                 95                 100                 105
Leu Glu Pro Arg Asn Gln Asn Phe Gln Glu Met Leu Arg Arg Leu
                110                 115                 120
Asn Thr Ser Ile Gln Glu Lys Leu Arg Val Gln Phe Ser Thr Asp
                125                 130                 135
Ser Arg Val Gln Lys Met Phe Glu Ile Leu Leu Asp Glu Asn Ser
                140                 145                 150
Glu Ala Asp Lys Arg Glu Lys Ala Ala Asn Asn Leu Ile Val Leu
                155                 160                 165
Gly Arg Glu Glu Ala Gly Ala Glu Lys Ile Phe Gln Asn Asn Gly
                170                 175                 180
Val Ala Leu Leu Leu Gln Leu Leu Asp Thr Lys Lys Pro Glu Leu
                185                 190                 195
Val Leu Ala Ala Val Arg Thr Leu Ser Gly Met Cys Ser Gly His
                200                 205                 210
```

Gln Ala Arg Ala Thr Val Ile Leu His Ala Val Arg Ile Asp Arg
215 220 225
Ile Cys Ser Leu Met Ala Val Glu Asn Glu Glu Met Ser Leu Ala
230 235 240
Val Cys Asn Leu Leu Gln Ala Ile Ile Asp Ser Leu Ser Gly Glu
245 250 255
Asp Lys Arg Glu His Arg Gly Lys Glu Glu Ala Leu Val Leu Asp
260 265 270
Thr Lys Lys Asp Leu Lys Gln Ile Thr Ser His Leu Leu Asp Met
275 280 285
Leu Val Ser Lys Lys Val Ser Gly Gln Gly Arg Asp Gln Ala Leu
290 295 300
Asn Leu Leu Asn Lys Asn Val Pro Arg Lys Asp Leu Ala Ile His
305 310 315
Asp Asn Ser Arg Thr Ile Tyr Val Val Asp Asn Gly Leu Arg Lys
320 325 330
Ile Leu Lys Val Val Gly Gln Val Pro Asp Leu Pro Ser Cys Leu
335 340 345
Pro Leu Thr Asp Asn Thr Arg Met Leu Ala Ser Ile Leu Ile Asn
350 355 360
Lys Leu Tyr Asp Asp Leu Arg Cys Asp Pro Glu Arg Asp His Phe
365 370 375
Arg Lys Ile Cys Glu Glu Tyr Ile Thr Gly Lys Phe Asp Pro Gln
380 385 390
Asp Met Asp Lys Asn Leu Asn Ala Ile Gln Thr Val Ser Gly Ile
395 400 405
Leu Gln Gly Pro Phe Asp Leu Gly Asn Gln Leu Leu Gly Leu Lys
410 415 420
Gly Val Met Glu Met Met Val Ala Leu Cys Gly Ser Glu Arg Glu
425 430 435
Thr Asp Gln Leu Val Ala Val Glu Ala Leu Ile His Ala Ser Thr
440 445 450
Lys Leu Ser Arg Ala Thr Phe Ile Ile Thr Asn Gly Val Ser Leu
455 460 465
Leu Lys Gln Ile Tyr Lys Thr Thr Lys Asn Glu Lys Ile Lys Ile
470 475 480
Arg Thr Leu Val Gly Leu Cys Lys Leu Gly Ser Ala Gly Gly Thr
485 490 495
Asp Tyr Gly Leu Arg Gln Phe Ala Glu Gly Ser Thr Glu Lys Leu
500 505 510
Ala Lys Gln Cys Arg Lys Trp Leu Cys Asn Met Ser Ile Asp Thr
515 520 525
Arg Thr Arg Arg Trp Ala Val Glu Gly Leu Ala Tyr Leu Thr Leu
530 535 540
Asp Ala Asp Val Lys Asp Asp Phe Val Gln Asp Val Pro Ala Leu
545 550 555
Gln Ala Met Phe Glu Leu Ala Lys Thr Ser Asp Lys Thr Ile Leu
560 565 570
Tyr Ser Val Ala Thr Thr Leu Val Asn Cys Thr Asn Ser Tyr Asp
575 580 585
Val Lys Glu Val Ile Pro Glu Leu Val Gln Leu Ala Lys Phe Ser
590 595 600
Lys Gln His Val Pro Glu Glu His Pro Lys Asp Lys Lys Asp Phe
605 610 615
Ile Asp Met Arg Val Lys Arg Leu Leu Lys Ala Gly Val Ile Ser
620 625 630

Ala Leu Ala Cys Met Val Lys Ala Asp Ser Ala Ile Leu Thr Asp
635 640 645
Gln Thr Lys Glu Leu Leu Ala Arg Val Phe Leu Ala Leu Cys Asp
650 655 660
Asn Pro Lys Asp Arg Gly Thr Ile Val Ala Gln Gly Gly Gly Lys
665 670 675
Ala Leu Ile Pro Leu Ala Leu Glu Gly Thr Asp Val Gly Lys Val
680 685 690
Lys Ala Ala His Ala Leu Ala Lys Ile Ala Ala Val Ser Asn Pro
695 700 705
Asp Ile Ala Phe Pro Gly Glu Arg Val Tyr Glu Val Val Arg Pro
710 715 720
Leu Val Arg Leu Leu Asp Thr Gln Arg Asp Gly Leu Gln Asn Tyr
725 730 735
Glu Ala Leu Leu Gly Leu Thr Asn Leu Ser Gly Arg Ser Asp Lys
740 745 750
Leu Arg Gln Lys Ile Phe Lys Glu Arg Ala Leu Pro Asp Ile Glu
755 760 765
Asn Tyr Met Phe Glu Asn His Asp Gln Leu Arg Gln Ala Ala Thr
770 775 780
Glu Cys Met Cys Asn Met Val Leu His Lys Glu Val Gln Glu Arg
785 790 795
Phe Leu Ala Asp Gly Asn Asp Arg Leu Lys Leu Val Val Leu Leu
800 805 810
Cys Gly Glu Asp Asp Asp Lys Val Gln Asn Ala Ala Ala Gly Ala
815 820 825
Leu Ala Met Leu Thr Ala Ala His Lys Lys Leu Cys Leu Lys Met
830 835 840
Thr Gln Val Thr Thr Gln Trp Leu Glu Ile Leu Gln Arg Leu Cys
845 850 855
Leu His Asp Gln Leu Ser Val Gln His Arg Gly Leu Val Ile Ala
860 865 870
Tyr Asn Leu Leu Ala Ala Asp Ala Glu Leu Ala Lys Lys Leu Val
875 880 885
Glu Ser Glu Leu Leu Glu Ile Leu Thr Val Val Gly Lys Gln Glu
890 895 900
Pro Asp Glu Lys Lys Ala Glu Val Val Gln Thr Ala Arg Glu Cys
905 910 915
Leu Ile Lys Cys Met Asp Tyr Gly Phe Ile Lys Pro Val Ser
920 925

<210> 14
<211> 530
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1234259CD1

<400> 14
Met Met Ser Glu His Asp Leu Ala Asp Val Val Gln Ile Ala Val
1 5 10 15
Glu Asp Leu Ser Pro Asp His Pro Val Val Leu Glu Asn His Val
20 25 30
Val Thr Asp Glu Asp Glu Pro Ala Leu Lys Arg Gln Arg Leu Glu

```
                35                  40                  45
Ile Asn Cys Gln Asp Pro Ser Ile Lys Ser Phe Leu Tyr Ser Ile
                50                  55                  60
Asn Gln Thr Ile Cys Leu Arg Leu Asp Ser Ile Glu Ala Lys Leu
                65                  70                  75
Gln Ala Leu Glu Ala Thr Cys Lys Ser Leu Glu Glu Lys Leu Asp
                80                  85                  90
Leu Val Thr Asn Lys Gln His Ser Pro Ile Gln Val Pro Met Val
                95                 100                 105
Ala Gly Ser Pro Leu Gly Ala Thr Gln Thr Cys Asn Lys Val Arg
               110                 115                 120
Cys Val Val Pro Gln Thr Thr Val Ile Leu Asn Asn Asp Arg Gln
               125                 130                 135
Asn Ala Ile Val Ala Lys Met Glu Asp Pro Leu Ser Asn Arg Ala
               140                 145                 150
Pro Asp Ser Leu Glu Asn Val Ile Ser Asn Ala Val Pro Gly Arg
               155                 160                 165
Arg Gln Asn Thr Ile Val Val Lys Val Pro Gly Gln Glu Asp Ser
               170                 175                 180
His His Glu Asp Gly Glu Ser Gly Ser Glu Ala Ser Asp Ser Val
               185                 190                 195
Ser Ser Cys Gly Gln Ala Gly Ser Gln Ser Ile Gly Ser Asn Val
               200                 205                 210
Thr Leu Ile Thr Leu Asn Ser Glu Glu Asp Tyr Pro Asn Gly Thr
               215                 220                 225
Trp Leu Gly Asp Glu Asn Asn Pro Glu Met Arg Val Arg Cys Ala
               230                 235                 240
Ile Ile Pro Ser Asp Met Leu His Ile Ser Thr Asn Cys Arg Thr
               245                 250                 255
Ala Glu Lys Met Ala Leu Thr Leu Leu Asp Tyr Leu Phe His Arg
               260                 265                 270
Glu Val Gln Ala Val Ser Asn Leu Ser Gly Gln Gly Lys His Gly
               275                 280                 285
Lys Lys Gln Leu Asp Pro Leu Thr Ile Tyr Gly Ile Arg Cys His
               290                 295                 300
Leu Phe Tyr Lys Phe Gly Ile Thr Glu Ser Asp Trp Tyr Arg Ile
               305                 310                 315
Lys Gln Ser Ile Asp Ser Lys Cys Arg Thr Ala Trp Arg Arg Lys
               320                 325                 330
Gln Arg Gly Gln Ser Leu Ala Val Lys Ser Phe Ser Arg Arg Thr
               335                 340                 345
Pro Asn Ser Ser Ser Tyr Cys Pro Ser Glu Pro Met Met Ser Thr
               350                 355                 360
Pro Pro Pro Ala Ser Glu Leu Pro Gln Pro Gln Pro Gln Pro Gln
               365                 370                 375
Ala Leu His Tyr Ala Leu Ala Asn Ala Gln Gln Val Gln Ile His
               380                 385                 390
Gln Ile Gly Glu Asp Gly Gln Val Gln Val Ile Pro Gln Gly His
               395                 400                 405
Leu His Ile Ala Gln Val Pro Gln Gly Glu Gln Val Gln Ile Thr
               410                 415                 420
Gln Asp Ser Glu Gly Asn Leu Gln Ile His His Val Gly Gln Asp
               425                 430                 435
Gly Gln Leu Leu Glu Ala Thr Arg Ile Pro Cys Leu Leu Ala Pro
               440                 445                 450
Ser Val Phe Lys Ala Ser Ser Gly Gln Val Leu Gln Gly Ala Gln
```

```
                455                 460                 465
Leu Ile Ala Val Ala Ser Ser Asp Pro Ala Ala Ala Gly Val Asp
                470                 475                 480
Gly Ser Pro Leu Gln Gly Ser Asp Ile Gln Val Gln Tyr Val Gln
                485                 490                 495
Leu Ala Pro Val Ser Asp His Thr Ala Gly Ala Gln Thr Ala Glu
                500                 505                 510
Ala Leu Gln Pro Thr Leu Gln Pro Glu Met Gln Leu Glu His Gly
                515                 520                 525
Ala Ile Gln Ile Gln
                530

<210> 15
<211> 821
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1440608CD1

<400> 15
Met Ala Lys Phe Ala Leu Asn Gln Asn Leu Pro Asp Leu Gly Gly
 1               5                  10                  15
Pro Arg Leu Cys Pro Val Pro Ala Ala Gly Gly Ala Arg Ser Pro
                20                  25                  30
Ser Ser Pro Tyr Ser Val Glu Thr Pro Tyr Gly Phe His Leu Asp
                35                  40                  45
Leu Asp Phe Leu Lys Tyr Ile Glu Glu Leu Glu Arg Gly Pro Ala
                50                  55                  60
Ala Arg Arg Ala Pro Gly Pro Pro Thr Ser Arg Arg Pro Arg Ala
                65                  70                  75
Pro Arg Pro Gly Leu Ala Gly Ala Arg Ser Pro Gly Ala Trp Thr
                80                  85                  90
Ser Ser Glu Ser Leu Ala Ser Asp Asp Gly Gly Ala Pro Gly Ile
                95                  100                 105
Leu Ser Gln Gly Ala Pro Ser Gly Leu Leu Met Gln Pro Leu Ser
                110                 115                 120
Pro Arg Ala Pro Val Arg Asn Pro Arg Val Glu His Thr Leu Arg
                125                 130                 135
Glu Thr Ser Arg Arg Leu Glu Leu Ala Gln Thr His Glu Arg Ala
                140                 145                 150
Pro Ser Pro Gly Arg Gly Val Pro Arg Ser Pro Arg Gly Ser Gly
                155                 160                 165
Arg Ser Ser Pro Ala Pro Asn Leu Ala Pro Ala Ser Pro Gly Pro
                170                 175                 180
Ala Gln Leu Gln Leu Val Arg Glu Gln Met Ala Ala Ala Leu Arg
                185                 190                 195
Arg Leu Arg Glu Leu Glu Asp Gln Ala Arg Thr Leu Pro Glu Leu
                200                 205                 210
Gln Glu Gln Val Arg Ala Leu Arg Ala Glu Lys Ala Arg Leu Leu
                215                 220                 225
Ala Gly Arg Ala Gln Pro Glu Pro Asp Gly Glu Ala Glu Thr Arg
                230                 235                 240
Pro Asp Lys Leu Ala Gln Leu Arg Arg Leu Thr Glu Arg Leu Ala
                245                 250                 255
```

Thr Ser Glu Arg Gly Gly Arg Ala Arg Ala Ser Pro Arg Ala Asp
260 265 270
Ser Pro Asp Gly Leu Ala Ala Gly Arg Ser Glu Gly Ala Leu Gln
275 280 285
Val Leu Asp Gly Glu Val Gly Ser Leu Asp Gly Thr Pro Gln Thr
290 295 300
Arg Glu Val Ala Ala Glu Ala Val Pro Glu Thr Arg Glu Ala Gly
305 310 315
Ala Gln Ala Val Pro Glu Thr Arg Glu Ala Gly Val Glu Ala Ala
320 325 330
Pro Glu Thr Val Glu Ala Asp Ala Trp Val Thr Glu Ala Leu Leu
335 340 345
Gly Leu Pro Ala Ala Ala Glu Arg Glu Leu Glu Leu Leu Arg Ala
350 355 360
Ser Leu Glu His Gln Arg Gly Val Ser Glu Leu Leu Arg Gly Arg
365 370 375
Leu Arg Glu Leu Glu Glu Ala Arg Glu Ala Ala Glu Glu Ala Ala
380 385 390
Ala Gly Ala Arg Ala Gln Leu Arg Glu Ala Thr Thr Gln Thr Pro
395 400 405
Trp Ser Cys Ala Glu Lys Ala Ala Gln Thr Glu Ser Pro Ala Glu
410 415 420
Ala Pro Ser Leu Thr Gln Glu Ser Ser Pro Gly Ser Met Asp Gly
425 430 435
Asp Arg Ala Val Ala Pro Ala Gly Ile Leu Lys Ser Ile Met Lys
440 445 450
Lys Arg Asp Gly Thr Pro Gly Ala Gln Pro Ser Ser Gly Pro Lys
455 460 465
Ser Leu Gln Phe Val Gly Val Leu Asn Gly Glu Tyr Glu Ser Ser
470 475 480
Ser Ser Glu Asp Ala Ser Asp Ser Asp Gly Asp Ser Glu Asn Gly
485 490 495
Gly Ala Glu Pro Pro Gly Ser Ser Ser Gly Ser Gly Asp Asp Ser
500 505 510
Gly Gly Gly Ser Asp Ser Gly Thr Pro Gly Pro Pro Ser Gly Gly
515 520 525
Asp Ile Arg Asp Pro Glu Pro Glu Ala Glu Ala Glu Pro Gln Gln
530 535 540
Val Ala Gln Gly Arg Cys Glu Leu Ser Pro Arg Leu Arg Glu Ala
545 550 555
Cys Val Ala Leu Gln Arg Gln Leu Ser Arg Pro Arg Gly Val Ala
560 565 570
Ser Asp Gly Gly Ala Val Arg Leu Val Ala Gln Glu Trp Phe Arg
575 580 585
Val Ser Ser Gln Arg Arg Ser Gln Ala Glu Pro Val Ala Arg Met
590 595 600
Leu Glu Gly Val Arg Arg Leu Gly Pro Glu Leu Leu Ala His Val
605 610 615
Val Asn Leu Ala Asp Gly Asn Gly Asn Thr Ala Leu His Tyr Ser
620 625 630
Val Ser His Gly Asn Leu Ala Ile Ala Ser Leu Leu Leu Asp Thr
635 640 645
Gly Ala Cys Glu Val Asn Arg Gln Asn Arg Ala Gly Tyr Ser Ala
650 655 660
Leu Met Leu Ala Ala Leu Thr Ser Val Arg Gln Glu Glu Glu Asp
665 670 675

```
Met Ala Val Val Gln Arg Leu Phe Cys Met Gly Asp Val Asn Ala
                680                 685                 690
Lys Ala Ser Gln Thr Gly Gln Thr Ala Leu Met Leu Ala Ile Ser
                695                 700                 705
His Gly Arg Gln Asp Met Val Ala Thr Leu Leu Ala Cys Gly Ala
                710                 715                 720
Asp Val Asn Ala Gln Asp Ala Asp Gly Ala Thr Ala Leu Met Cys
                725                 730                 735
Ala Ser Glu Tyr Gly Arg Leu Asp Thr Val Arg Leu Leu Leu Thr
                740                 745                 750
Gln Pro Gly Cys Asp Pro Ala Ile Leu Asp Asn Glu Gly Thr Ser
                755                 760                 765
Ala Leu Ala Ile Ala Leu Glu Ala Glu Gln Asp Glu Val Ala Ala
                770                 775                 780
Leu Leu His Ala His Leu Ser Ser Gly Gln Pro Asp Thr Gln Ser
                785                 790                 795
Glu Ser Pro Pro Gly Ser Gln Thr Ala Thr Pro Gly Glu Gly Glu
                800                 805                 810
Cys Gly Asp Asn Gly Glu Asn Pro Gln Val Gln
                815                 820

<210> 16
<211> 1003
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3413610CD1

<400> 16
Met Ala Arg Arg Gly Lys Lys Pro Val Val Arg Thr Leu Glu Asp
  1               5                  10                  15
Leu Thr Leu Asp Ser Gly Tyr Gly Gly Ala Ala Asp Ser Val Arg
                 20                  25                  30
Ser Ser Asn Leu Ser Leu Cys Cys Ser Asp Ser His Pro Ala Ser
                 35                  40                  45
Pro Tyr Gly Gly Ser Cys Trp Pro Pro Leu Ala Asp Ser Met His
                 50                  55                  60
Ser Arg His Asn Ser Phe Asp Thr Val Asn Thr Ala Leu Val Glu
                 65                  70                  75
Asp Ser Glu Gly Leu Asp Cys Ala Gly Gln His Cys Ser Arg Leu
                 80                  85                  90
Leu Pro Asp Leu Asp Glu Val Pro Trp Thr Leu Gln Glu Leu Glu
                 95                 100                 105
Ala Leu Leu Leu Arg Ser Arg Asp Pro Arg Ala Gly Pro Ala Val
                110                 115                 120
Pro Gly Gly Leu Pro Lys Asp Ala Leu Ala Lys Leu Ser Thr Leu
                125                 130                 135
Val Ser Arg Ala Leu Val Arg Ile Ala Lys Glu Ala Gln Arg Leu
                140                 145                 150
Ser Leu Arg Phe Ala Lys Cys Thr Lys Tyr Glu Ile Gln Ser Ala
                155                 160                 165
Met Glu Ile Val Leu Ser Trp Gly Leu Ala Ala His Cys Thr Ala
                170                 175                 180
Ala Ala Leu Ala Ala Leu Ser Leu Tyr Asn Met Ser Ser Ala Gly
```

185 190 195
Gly Asp Arg Leu Gly Arg Gly Lys Ser Ala Arg Cys Gly Leu Thr
200 205 210
Phe Ser Val Gly Arg Val Tyr Arg Trp Met Val Asp Ser Arg Val
215 220 225
Ala Leu Arg Ile His Glu His Ala Ala Ile Tyr Leu Thr Ala Cys
230 235 240
Met Glu Ser Leu Phe Arg Asp Ile Tyr Ser Arg Val Val Ala Ser
245 250 255
Gly Val Pro Arg Ser Cys Ser Gly Pro Gly Ser Gly Ser Gly Ser
260 265 270
Gly Pro Gly Pro Ser Ser Gly Pro Gly Ala Ala Pro Ala Ala Asp
275 280 285
Lys Glu Arg Glu Ala Pro Gly Gly Gly Ala Ala Ser Gly Gly Ala
290 295 300
Cys Ser Ala Ala Ser Ser Ala Ser Gly Gly Ser Ser Cys Cys Ala
305 310 315
Pro Pro Ala Ala Ala Ala Ala Ala Val Pro Pro Thr Ala Ala Ala
320 325 330
Asn His His His His His His His Ala Leu His Glu Ala Pro Lys
335 340 345
Phe Thr Val Glu Thr Leu Glu His Thr Val Asn Asn Asp Ser Glu
350 355 360
Ile Trp Gly Leu Leu Gln Pro Tyr Gln His Leu Ile Cys Gly Lys
365 370 375
Asn Ala Ser Gly Asp Leu Val Ser Arg Ala Met His His Leu Gln
380 385 390
Pro Leu Gln Val Glu Arg Pro Phe Leu Val Leu Pro Pro Leu Met
395 400 405
Glu Trp Ile Arg Val Ala Val Ala His Ala Gly His Arg Arg Ser
410 415 420
Phe Ser Met Asp Ser Asp Asp Val Arg Gln Ala Ala Arg Leu Leu
425 430 435
Leu Pro Gly Val Asp Cys Glu Pro Arg Gln Leu Arg Ala Asp Asp
440 445 450
Cys Phe Cys Ala Ser Arg Lys Leu Asp Ala Val Ala Ile Glu Ala
455 460 465
Lys Phe Lys Gln Asp Leu Gly Phe Arg Met Leu Asn Cys Gly Arg
470 475 480
Thr Asp Leu Val Lys Gln Ala Val Ser Leu Leu Gly Pro Asp Gly
485 490 495
Ile Asn Thr Met Ser Glu Gln Gly Met Thr Pro Leu Met Tyr Ala
500 505 510
Cys Val Arg Gly Asp Glu Ala Met Val Gln Met Leu Leu Asp Ala
515 520 525
Gly Ala Asp Leu Asn Val Glu Val Val Ser Thr Pro His Lys Tyr
530 535 540
Pro Ser Val His Pro Glu Thr Arg His Trp Thr Ala Leu Thr Phe
545 550 555
Ala Val Leu His Gly His Ile Pro Val Val Gln Leu Leu Leu Asp
560 565 570
Ala Gly Ala Lys Val Glu Gly Ser Val Glu His Gly Glu Glu Asn
575 580 585
Tyr Ser Glu Thr Pro Leu Gln Leu Ala Ala Ala Val Gly Asn Phe
590 595 600
Glu Leu Val Ser Leu Leu Leu Glu Arg Gly Ala Asp Pro Leu Ile

```
                605                 610                 615
Gly Thr Met Tyr Arg Asn Gly Ile Ser Thr Thr Pro Gln Gly Asp
                620                 625                 630
Met Asn Ser Phe Ser Gln Ala Ala Ala His Gly His Arg Asn Val
                635                 640                 645
Phe Arg Lys Leu Leu Ala Gln Pro Glu Lys Glu Lys Ser Asp Ile
                650                 655                 660
Leu Ser Leu Glu Glu Ile Leu Ala Glu Gly Thr Asp Leu Ala Glu
                665                 670                 675
Thr Ala Pro Pro Pro Leu Cys Ala Ser Arg Asn Ser Lys Ala Lys
                680                 685                 690
Leu Arg Ala Leu Arg Glu Ala Met Tyr His Ser Ala Glu His Gly
                695                 700                 705
Tyr Val Asp Val Thr Ile Asp Ile Arg Ser Ile Gly Val Pro Trp
                710                 715                 720
Thr Leu His Thr Trp Leu Glu Ser Leu Arg Ile Ala Phe Gln Gln
                725                 730                 735
His Arg Arg Pro Leu Ile Gln Cys Leu Leu Lys Glu Phe Lys Thr
                740                 745                 750
Ile Gln Glu Glu Glu Tyr Thr Glu Glu Leu Val Thr Gln Gly Leu
                755                 760                 765
Pro Leu Met Phe Glu Ile Leu Lys Ala Ser Lys Asn Glu Val Ile
                770                 775                 780
Ser Gln Gln Leu Cys Val Ile Phe Thr His Cys Tyr Gly Pro Tyr
                785                 790                 795
Pro Ile Pro Lys Leu Thr Glu Ile Lys Arg Lys Gln Thr Ser Arg
                800                 805                 810
Leu Asp Pro His Phe Leu Asn Asn Lys Glu Met Ser Asp Val Thr
                815                 820                 825
Phe Leu Val Glu Gly Arg Pro Phe Tyr Ala His Lys Val Leu Leu
                830                 835                 840
Phe Thr Ala Ser Pro Arg Phe Lys Ala Leu Leu Ser Ser Lys Pro
                845                 850                 855
Thr Asn Asp Gly Thr Cys Ile Glu Ile Gly Tyr Val Lys Tyr Ser
                860                 865                 870
Ile Phe Gln Leu Val Met Gln Tyr Leu Tyr Tyr Gly Gly Pro Glu
                875                 880                 885
Ser Leu Leu Ile Lys Asn Asn Glu Ile Met Glu Leu Leu Ser Ala
                890                 895                 900
Ala Lys Phe Phe Gln Leu Glu Ala Leu Gln Arg His Cys Glu Ile
                905                 910                 915
Ile Cys Ala Lys Ser Ile Asn Thr Asp Asn Cys Val Asp Ile Tyr
                920                 925                 930
Asn His Ala Lys Phe Leu Gly Val Thr Glu Leu Ser Ala Tyr Cys
                935                 940                 945
Glu Gly Tyr Phe Leu Lys Asn Met Met Val Leu Ile Glu Asn Glu
                950                 955                 960
Ala Phe Lys Gln Leu Leu Tyr Asp Lys Asn Gly Glu Gly Thr Gly
                965                 970                 975
Gln Asp Val Leu Gln Asp Leu Gln Arg Thr Leu Ala Ile Arg Ile
                980                 985                 990
Gln Ser Ile His Leu Ser Ser Ser Lys Gly Ser Val Val
                995                1000
```

<210> 17
<211> 888

<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3276394CD1

<400> 17
Met Asp Glu Ser Ala Leu Leu Asp Leu Leu Glu Cys Pro Val Cys
1 5 10 15
Leu Glu Arg Leu Asp Ala Ser Ala Lys Val Leu Pro Cys Gln His
20 25 30
Thr Phe Cys Lys Arg Cys Leu Leu Gly Ile Val Gly Ser Arg Asn
35 40 45
Glu Leu Arg Cys Pro Glu Cys Arg Thr Leu Val Gly Ser Gly Val
50 55 60
Glu Glu Leu Pro Ser Asn Ile Leu Leu Val Arg Leu Leu Asp Gly
65 70 75
Ile Lys Gln Arg Pro Trp Lys Pro Gly Pro Gly Gly Gly Ser Gly
80 85 90
Thr Asn Cys Thr Asn Ala Leu Arg Ser Gln Ser Ser Thr Val Ala
95 100 105
Asn Cys Ser Ser Lys Asp Leu Gln Ser Ser Gln Gly Gly Gln Gln
110 115 120
Pro Arg Val Gln Ser Trp Ser Pro Pro Val Arg Gly Ile Pro Gln
125 130 135
Leu Pro Cys Ala Lys Ala Leu Tyr Asn Tyr Glu Gly Lys Glu Pro
140 145 150
Gly Asp Leu Lys Phe Ser Lys Gly Asp Ile Ile Ile Leu Arg Arg
155 160 165
Gln Val Asp Glu Asn Trp Tyr His Gly Glu Val Asn Gly Ile His
170 175 180
Gly Phe Phe Pro Thr Asn Phe Val Gln Ile Ile Lys Pro Leu Pro
185 190 195
Gln Pro Pro Ser Gln Cys Lys Ala Leu Tyr Asp Phe Glu Val Lys
200 205 210
Asp Lys Glu Ala Asp Lys Asp Cys Leu Pro Phe Ala Lys Asp Asp
215 220 225
Val Leu Thr Val Ile Arg Arg Val Asp Glu Asn Trp Ala Glu Gly
230 235 240
Met Leu Ala Asp Lys Ile Gly Ile Phe Pro Ile Ser Tyr Val Glu
245 250 255
Phe Asn Ser Ala Ala Lys Gln Leu Ile Glu Trp Asp Lys Pro Pro
260 265 270
Val Pro Gly Val Asp Ala Gly Glu Cys Ser Ser Ala Ala Ala Gln
275 280 285
Ser Ser Thr Ala Pro Lys His Ser Asp Thr Lys Lys Asn Thr Lys
290 295 300
Lys Arg His Ser Phe Thr Ser Leu Thr Met Ala Asn Lys Ser Ser
305 310 315
Gln Ala Ser Gln Asn Arg His Ser Met Glu Ile Ser Pro Pro Val
320 325 330
Leu Ile Ser Ser Ser Asn Pro Thr Ala Ala Ala Arg Ile Ser Glu
335 340 345
Leu Ser Gly Leu Ser Cys Ser Ala Pro Ser Gln Val His Ile Ser
350 355 360

Thr Thr Gly Leu Ile Val Thr Pro Pro Pro Ser Ser Pro Val Thr
365 370 375
Thr Gly Pro Ser Phe Thr Phe Pro Ser Asp Val Pro Tyr Gln Ala
380 385 390
Ala Leu Gly Thr Leu Asn Pro Pro Leu Pro Pro Pro Pro Leu Leu
395 400 405
Ala Ala Thr Val Leu Ala Ser Thr Pro Pro Gly Ala Thr Ala Ala
410 415 420
Ala Ala Ala Ala Gly Met Gly Pro Arg Pro Met Ala Gly Ser Thr
425 430 435
Asp Gln Ile Ala His Leu Arg Pro Gln Thr Arg Pro Ser Val Tyr
440 445 450
Val Ala Ile Tyr Pro Tyr Thr Pro Arg Lys Glu Asp Glu Leu Glu
455 460 465
Leu Arg Lys Gly Glu Met Phe Leu Val Phe Glu Arg Cys Gln Asp
470 475 480
Gly Trp Phe Lys Gly Thr Ser Met His Thr Ser Lys Ile Gly Val
485 490 495
Phe Pro Gly Asn Tyr Val Ala Pro Val Thr Arg Ala Val Thr Asn
500 505 510
Ala Ser Gln Ala Lys Val Pro Met Ser Thr Ala Gly Gln Thr Ser
515 520 525
Arg Gly Val Thr Met Val Ser Pro Ser Thr Ala Gly Gly Pro Ala
530 535 540
Gln Lys Leu Gln Gly Asn Gly Val Ala Gly Ser Pro Ser Val Val
545 550 555
Pro Ala Ala Val Val Ser Ala Ala His Ile Gln Thr Ser Pro Gln
560 565 570
Ala Lys Val Leu Leu His Met Thr Gly Gln Met Thr Val Asn Gln
575 580 585
Ala Arg Asn Ala Val Arg Thr Val Ala Ala His Asn Gln Glu Arg
590 595 600
Pro Thr Ala Ala Val Thr Pro Ile Gln Val Gln Asn Ala Ala Gly
605 610 615
Leu Ser Pro Ala Ser Val Gly Leu Ser His His Ser Leu Ala Ser
620 625 630
Pro Gln Pro Ala Pro Leu Met Pro Gly Ser Ala Thr His Thr Ala
635 640 645
Ala Ile Ser Ile Ser Arg Ala Ser Ala Pro Leu Ala Cys Ala Ala
650 655 660
Ala Ala Pro Leu Thr Ser Pro Ser Ile Thr Ser Ala Ser Leu Glu
665 670 675
Ala Glu Pro Ser Gly Arg Ile Val Thr Val Leu Pro Gly Leu Pro
680 685 690
Thr Ser Pro Asp Ser Ala Ser Ser Ala Cys Gly Asn Ser Ser Ala
695 700 705
Thr Lys Pro Asp Lys Asp Ser Lys Lys Glu Lys Lys Gly Leu Leu
710 715 720
Lys Leu Leu Ser Gly Ala Ser Thr Lys Arg Lys Pro Arg Val Ser
725 730 735
Pro Pro Ala Ser Pro Thr Leu Glu Val Glu Leu Gly Ser Ala Glu
740 745 750
Leu Pro Leu Gln Gly Ala Val Gly Pro Glu Leu Pro Pro Gly Gly
755 760 765
Gly His Gly Arg Ala Gly Ser Cys Pro Val Asp Gly Asp Gly Pro
770 775 780

Val Thr Thr Ala Val Ala Gly Ala Ala Leu Ala Gln Asp Ala Phe
785 790 795
His Arg Lys Ala Ser Ser Leu Asp Ser Ala Val Pro Ile Ala Pro
800 805 810
Pro Pro Arg Gln Ala Cys Ser Ser Leu Gly Pro Val Leu Asn Glu
815 820 825
Ser Arg Pro Val Val Cys Glu Arg His Arg Val Val Val Ser Tyr
830 835 840
Pro Pro Gln Ser Glu Ala Glu Leu Glu Leu Lys Glu Gly Asp Ile
845 850 855
Val Phe Val His Lys Lys Arg Glu Asp Gly Trp Phe Lys Gly Thr
860 865 870
Leu Gln Arg Asn Gly Lys Thr Gly Leu Phe Pro Gly Ser Phe Val
875 880 885
Glu Asn Ile

<210> 18
<211> 283
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7602049CD1

<400> 18
Met Ser Tyr Ser Val Thr Leu Thr Gly Pro Gly Pro Trp Gly Phe
1 5 10 15
Arg Leu Gln Gly Gly Lys Asp Phe Asn Met Pro Leu Thr Ile Ser
20 25 30
Arg Ile Thr Pro Gly Ser Lys Ala Ala Gln Ser Gln Leu Ser Gln
35 40 45
Gly Asp Leu Val Val Ala Ile Asp Gly Val Asn Thr Asp Thr Met
50 55 60
Thr His Leu Glu Ala Gln Asn Lys Ile Lys Ser Ala Ser Tyr Asn
65 70 75
Leu Ser Leu Thr Leu Gln Lys Ser Lys Arg Pro Ile Pro Ile Ser
80 85 90
Thr Thr Ala Pro Pro Val Gln Thr Pro Leu Pro Val Ile Pro His
95 100 105
Gln Lys Val Val Val Asn Ser Pro Ala Asn Ala Asp Tyr Gln Glu
110 115 120
Arg Phe Asn Pro Ser Ala Leu Lys Asp Ser Ala Leu Ser Thr His
125 130 135
Lys Pro Ile Glu Val Lys Gly Leu Gly Gly Lys Ala Thr Ile Ile
140 145 150
His Ala Gln Tyr Asn Thr Pro Ile Ser Met Tyr Ser Gln Asp Ala
155 160 165
Ile Met Asp Ala Ile Ala Gly Gln Ala Gln Ala Gln Gly Ser Asp
170 175 180
Phe Ser Gly Ser Leu Pro Ile Lys Asp Leu Ala Val Asp Ser Ala
185 190 195
Ser Pro Val Tyr Gln Ala Val Ile Lys Ser Gln Asn Lys Pro Glu
200 205 210
Asp Glu Ala Asp Glu Trp Ala Arg Arg Ser Ser Asn Leu Gln Ser

```
                215                 220                 225
Arg Ser Phe Arg Ile Leu Ala Gln Met Thr Gly Thr Glu Phe Met
                230                 235                 240
Gln Asp Pro Asp Glu Glu Ala Leu Arg Arg Ser Arg Glu Arg Phe
                245                 250                 255
Glu Thr Glu Arg Asn Ser Pro Arg Phe Ala Lys Leu Arg Asn Trp
                260                 265                 270
His His Gly Leu Ser Ala Gln Ile Leu Asn Val Lys Ser
                275                 280
```

<210> 19
<211> 1830
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5566074CB1

<400> 19

```
atgtacacct tcgtggtacg cgatgagaac agcagcgtct acgccgaggt ctcccggctg 60
ctcctcgcca ccggccactg gaagaggctg cggcgagaca accccagatt caacctgatg 120
ctgggagaga ggaatcggct gcccttcggg agactgggtc acgagcccgg gctggtacag 180
ttggtgaatt actacagggg tgctgacaaa ctgtgtcgca aagcttcttt agtgaagcta 240
atcaagacaa gccctgaact ggctgagtcc tgcacatggt tccctgaatc ttatgtgatt 300
tatccaacca atctcaagac tccagttgct ccagcacaga atggaattca gccaccaatc 360
agtaactcaa ggacagatga aagagaattc tttctcgcct cttataacag aaagaaagag 420
gatggagagg gcaacgtttg gattgcaaag tcatcagccg gtgccaaagg tgaaggcatt 480
ctcatctcct cagaggcttc agagcttctc gatttcatag acaaccaggg ccaagtgcac 540
gtgatccaga aatatcttga gcaccctctg ctgcttgagc caggtcatcg caagtttgac 600
attcgaagct gggtcttggt ggatcatcag tataatatct acctctatag agagggtgtg 660
cttcggactg cttcagaacc atatcatgtt gataatttcc aagacaaaac ctgccatttg 720
accaatcact gcattcaaaa agagtattca aagaactacg ggaagtatga agaaggaaat 780
gaaatgttct tcaaggagtt caatcagtac ctaacaagtg ctttgaacat taccctagaa 840
agtagtatct tactacaaat caaacatata ataaggaact gcctcctgag cgtggagcct 900
gccattagca ccaagcacct cccttaccag agcttccagc tcttcggctt tgacttcatg 960
gtcgatgagg agctgaaggt gtggctcatt gaggtcaacg gtgcccctgc atgtgctcag 1020
aagctctatg cagaactgtg ccaaggcatc gtggacatag ccatttccag tgtcttccca 1080
cccccagatg tggagcaacc tcagacccag ccagctgcct tcatcaagct gtgacagagg 1140
gcactccctg ctgccttgga aaaagcacgg ggtcctgctc cagggaatgg tgaaatgact 1200
ggattgctct ttatccagcc cacagcaggg gaaagaaagg caactcgcaa agatgagatg 1260
gaagaaggca cgtgagcaga ggaggcagct cccaaagaga gggctgctca gggggcttcc 1320
caggtgtagc tctcagcagt gctgttgaga cttttgaaaa caactttggt acacaaaggc 1380
agctttgtga gcagagctcc ttcccctctc cccgggaacg gcagggcact gggacctctg 1440
gtcggtgcct cccacccact gcagccctag tgccttagct ccatgcccgg ctgcagcccc 1500
actgctctgg actatggatt ggacgtcaga gcatattgga ggttgcctgt gtgttcccca 1560
cccatccctt cggtaacact ctgccacact aagctctgta caagcatgca ccaacagtcc 1620
ttagttttgt gctgtgcact ggcctctcgg caaaggtggt ttccctcatc accttcctga 1680
tggtgtttgg tcagtcacct gtcagggttt gtgcgggttg ggccccaaaa cagcatatgc 1740
tgctctaagt ctgctctctg catgttttag aaacaaagtg gcaagtctgc cctgaacctg 1800
taagcatcaa ataagcatga gagagaaaaa                                  1830
```

<210> 20
<211> 2795
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5679814CB1

<400> 20
ggaaaaactc tctgctcgtc atcaaggcag catcatcatc gttattgatt ctatagatca 60
agttcagcaa gttgaaaaac acatgaaatg gctgatagat ccactgccag tgaatgtaag 120
agtaattgtt tctgtgaatg tagaaacatg ccctccagca tggaggttgt ggcctacact 180
tcatcttgat cccttaagtc caaaagatgc aaaatctatt ataattgcag aatgccactc 240
tgtagacatt aaattgagta aagagcagga gaagaagcta gaacgacact gtcgttctgc 300
tacaacctgc aatgcccttt atgtcaccct tttcggcaaa atgatcgcgc gtgctgggag 360
agcaggcaat ttagataaaa tccttcatca gtgtttccag tgtcaagata ctctttcatt 420
atatagactt gttctgcact ctatccggga gtccatggca aatgatgtgg ataaagagct 480
aatgaagcag atcctctgcc ttgtcaatgt tagtcacaat ggtgtgagtg aatcagaact 540
gatggaactc tatcctgaga tgtcctggac tttcttgacc tcccttattc acagtttata 600
caaaatgtgt ttgttgactt atggatgtgg cttgcttagg tttcaacatc tgcaggcttg 660
ggaaacagtg agattggagt acctggaagg ccccactgtt acttcttcat acaggcaaaa 720
gctaatcaac tatttcacct tgcagctaag tcaggacaga gtgacttgga gaagtgcaga 780
tgaactcccg tggctttttc agcagcaggg aagtaaacag aagctgcatg attgccttct 840
taatctcttt gtgtctcaaa acctttataa aaggggacac tttgctgagt tgctgagtta 900
ttggcagttt gttggcaaag acaaaagtgc aatggcaaca gaatacttcg attcattgaa 960
gcagtatgag aaaaactgcg aaggcgagga caacatgagt tgcttagctg atctttatga 1020
aaccttgggg cgatttctca aggatctagg ccttctcagt caggccatag tacctttgca 1080
gaggtcttta gagattcgag aaacagcttt agatcccgat cacccaagag tagcccagtc 1140
cctccaccaa ctagcaagtg tatacgtgca gtggaagaag tttggcaatg cagaacaact 1200
gtataaacag gcgttggaaa tctcagaaaa tgcttatggt gcggaccatc catatactgc 1260
tcgtgaactt gaagcacttg caactttgta ccagaaacaa aataaatatg aacaagctga 1320
acattttagg aaaaaatcct ttaaaattca tcagaaagct ataaagaaaa aaggcaactt 1380
gtacggattt gccctttttac gtagacgggc tttacagtta gaagagctta cattaggtaa 1440
ggacacacct gataatgctc ggaccctcaa tgaactgggt gttctctact atcttcaaaa 1500
taacctggaa acagctgacc agtttctgaa gcgttcctta gaaatgaggg agcgagttct 1560
aggaccagat caccctgact gtgctcagtc tttgaataat ctggcagctc tatgcaatga 1620
aaagaaacag tatgataaag cagaagaact ttatgaaaga gctttagata ttcggagacg 1680
tgcattagct cctgatcacc cttctttggc atatacggtg aagcatcttg ccatcttgta 1740
taagaaaatg ggaaacttg acaaagctgt acctttgtat gaattggctg ttgaaattcg 1800
acagaaatct tttggcccaa agcaccctag tgtagctact gccttggtga acttagctgt 1860
tctttatagc caaatgaaaa aacacgttga agctttgcca ttatatgaaa gagcattaaa 1920
gatttatgaa gatagcctgg gtcggatgca tcctcgagtt ggagaaacac tgaaaaattt 1980
agctgtgctt agctatgaag gaggagattt tgaaaaagct gctgaattat acaaaagggc 2040
aatggaaata aaagaagcag aaacatcact cttgggtgga aaagctcctt cacgccattc 2100
atcaagtgga gacacgttta gcttaaaaac agctcattct cctaatgttt tccttcagca 2160
aggacaaagg taatagcagc agttagaatt ctttgcaaat gtaccttaag acaaaataat 2220
taaacatttg gaacatttga atttgaaact ttaaaaaaat gttgtacgaa attttactac 2280
gtgtgattta actgctattt gtatgaagtt gtattggatt acattaagtt ggaattgtga 2340
ttatgtctgt tttagttgtt taaaagaatt ttcctattat atggtatcca aggatgtaga 2400
cacattagaa ttataagaag acatgaggag caaatcatga agagcggatt ggtctttgtt 2460
caacaagagc tggcagagta gttaagacaa ggagttcaaa aattccatga atcttggcca 2520
ggcatggtgg ctcatgcctg taatcccacc actttgggag ggtgatgcag gaggatcact 2580
tgaggccagg agttggagac cagcctggcc aacatggtga aaccctgtct ctactaaaaa 2640
tacaaaaaaa aattagctgg acctggtggc gcatgcctgt aatcctagct actcaggagg 2700
ctgaggcatg agaatcgctt gaactcagca agtggaggtt gctatgagct gagattgtgc 2760
cagttcattc cacactgggc aacagggtga gctga 2795

<210> 21
<211> 4436
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472735CB1

<400> 21
gccagatcgc cgcgcgaggg atggtgggca tcgaggtccc agcagcggac gagggaggtg 60
ccgccgtcgc ccaggatggg ctgggaatga agcgatgtag ccttttaaga gatttgctct 120
gacccatctg aagtccatat ggctctgtat gatgaagacc tcctgaaaaa tcctttctat 180
ctggctctgc aaaagtgccg ccctgacttg tgcagcaaag tggcccaaat ccatggcatt 240
gtcttagtac cctgcaaagg aagcctgtcg agcagcatcc agtctacttg tcagtttgag 300
tcctacattt tgatacctgt ggaagagcat tttcagacct taaatggaaa ggatgtcttt 360
attcaaggga acaggattaa attaggagct ggttttgcct gtcttctctc agtgcccatt 420
ctctttgaag aaactttcta caatgaaaaa gaagagagtt tcagcatcct gtgtatagcc 480
catcctttgg aaaagagaga gagttcagaa gagcctttgg caccctcaga tccctttcc 540
ctgaaaacca ttgaagatgt gagagagttc ttgggaagac actccgagcg atttgacagg 600
aacatcgcct ctttccatcg aacattccga gaatgcgaga gaaagagcct ccgtcaccac 660
atagactcag cgaatgctct ctacaccaaa tgcctccagc agcttctgag ggactctcac 720
ctgaaaatgc tcgccaagca ggaggcccag atgaacctga tgaagcaggc agtggagata 780
tacgtccatc atgaaattta caacctgatc tttaaatacg tggggaccat ggaggcaagt 840
gaggatgcgg cctttaacaa aatcacaaga agccttcaag atcttcagca gaaagatatt 900
ggtgtgaaac cggagttcag ctttaacata cctcgtgcca aaagagagct ggctcagctg 960
aacaaatgca cctccccaca gcagaagctt gtctgcttgc gaaaagtggt gcagctcatt 1020
acacagtctc caagccagag agtgaacctg gagaccatgt gtgctgatga tctgctatca 1080
gtcctgttat acttgcttgt gaaaacggag atccctaatt ggatggcaaa tttgagttac 1140
atcaaaaact tcaggtttag cagcttggca aaggatgaac tgggatactg cctgacctca 1200
ttcgaagctg ccattgaata tattcggcaa ggaagcctct ctgctaaacc ccctgagtct 1260
gagggatttg gagacaggct gttccttaag cagagaatga gcttactctc tcagatgact 1320
tcgtctccca ccgactgcct gtttaagcac attgcatcag gtaaccagaa agaagtggag 1380
agacttctga gccaagagga ccatgataaa gataccgtcc aaaagatgtg tcaccctctc 1440
tgcttctgcg atgactgtga gaaactcgtc tctgggaggt tgaatgatcc ctcagttgtc 1500
actccattct ccagagacga caggggggcac acccctctcc atgtggctgc tgtctgtggg 1560
caggcatccc tcatcgacct cctggtttcc aagggcgcca tggtaaatgc cacagactac 1620
catggagcca ctccgctcca cctggcctgt cagaagggct accagagcgt gacgctgctg 1680
ctgctgcact acaaggccag cgcggaagtg caggacaaca atgggaatac gccactccac 1740
ctggcctgca cttacggcca cgaggactgt gtgaaggctc tggtttacta cgacgtggag 1800
tcgtgcagac ttgacattgg caatgagaaa ggagacaccc ctctacacat tgctgcccgc 1860
tggggctacc aaggcgtcat agagacattg ctgcagaacg gagcgtccac cgagatccag 1920
aacagactga aggagacgcc cctcaagtgt gcattaaact caaagattct gtctgtaatg 1980
gaagcctatc acctgtcctt cgagaggagg cagaagtcgt ccgaggcccc tgtgcagtcc 2040
ccgcagcgct ccgtggactc catcagccaa gagtcctcca cttccagctt ctcctccatg 2100
tcagccagct caaggcagga ggagaccaag aaggactaca gagaggtaga aaaacttttg 2160
agagcagttg ctgatggaga tctagaaatg gtgcgttacc tgttggaatg gacagaggag 2220
gacctggagg atgcggagga cactgtcagt gcagcggacc ccgaattctg tcacccgttg 2280
tgccagtgcc ccaagtgtgc cccagctcag aagagaggctgg cgaaggttcc tgccagtggg 2340
cttggtgtga acgtgaccag ccaggacggc tcctccccgc tgcatgtcgc cgccctgcac 2400
ggccgggcgg acctcatccc cctcctgctg aagcacgggg ccaacgcagg tgccaggaac 2460
gcagaccaag ccgtcccgct ccacctggcc tgccagcagg gccactttca ggtggtgaag 2520
tgtctgttag attcgaatgc aaaacccaat aagaaggacc tcagtggaaa cacgcccctc 2580
atttacgcct gctccggtgg ccatcacgag cttgtggcac tgctgctaca gcacggggcc 2640
tccattaacg cttctaacaa taagggcaac acagcgctgc acgaggctgt gattgaaaag 2700
cacgtcttcg tggtagagct gcttctgctc cacggagcgt cagttcaggt gctgaacaag 2760
cggcagcgca cggctgtaga ctgtgctgaa cagaattcaa aaataatgga attgcttcag 2820
gtggtaccaa gctgtgttgc ttcattagat gatgtggctg aaactgaccg caaggagtat 2880
gtcactgtta agatcaggaa aaaatggaac tcaaaactgt atgatctacc agatgagcct 2940

```
tttacaagac agttttactt tgtccactca gctggtcagt ttaagggaaa gacttcaagg 3000
gagattatgg caagagatag aagtgtccct aatttaaccg aaggttcttt gcatgagcca 3060
gggaggcaaa gtgtcacact gagacagaat aacctgccag ctcagagtgg atctcatgct 3120
gctgagaaag gcaacagcga ctggccagag aggcctggac tgacacagac tggccctgga 3180
cacagacgga tgctgcggag acacacggta gaggatgcgg tcgtgtccca gggcccggag 3240
gctgctggcc ccctctccac tccccaagag gttagtgctt cccggtccta acaggaatga 3300
ggagttgttg aacccactgc taggaagcaa ggatgcaaca agatgatgct gagcgtgaac 3360
acatctgaga actaaatgtg cttccatgag actggcttga gaagtcttca gcaccaagtt 3420
cctgaaagct tttctgtggc aggaaagaat gcaacaaaaa agttaaccac caccatctct 3480
ctcctcttca aagctaatga atacaattga aacagacaaa aattccagta gcatccagat 3540
ccttaagcca gaggtgcatg cttcttttta agtatgaggg tttgttggtc acagtgggag 3600
aggtttcacc accgcattct gacctcctcc tcccaaaagg tgctaaacct ctctgacctg 3660
tgtacattca caaaccacag ctagaattcc tccacctagg attaagctgg agagaagtaa 3720
gtaatttagg tttcatggta ctgtagaggc caggctgaaa tgtcatatct gaaggaagaa 3780
agcagcagct ggacaatgtt tctttgcaaa gcaacactcg aaccaaaaga tgcctcaatc 3840
ccattttgat attcatttta gtgaaaggat gcatcagacc tgttccacat catgcacatg 3900
ggaaagggtg gttatcattt tccttctaac aagtaggtac agatattcgg ttactacacg 3960
tgcacctgta gcagtatttc tagaaacatc ccttttttgtt gagaacctcc cttgaatgtc 4020
tgtcacactc acacctgacg ggatggttac tggattagag agtagatttg gcacatcttt 4080
tcttagtctt ttgattcaaa ttcaaaactt aacagcacaa accaggtcag agttactttc 4140
ggttagaatt tattgccatt tattcctttt tataaatttc tatagattat actgttattt 4200
ttatgttatt ggcctagagc tacacgtata tgggtttgtc ctgagtccgt tttcaaatga 4260
ccttgtgata gggaaatggt tttgtccatg ttcttggaaa cacttgtgta tgtacagaag 4320
gaagggaggg attatttttc tacaaagtaa tttatgattt ctaattttct aatgtgcctt 4380
ggatatgtgc caaatgatgg aaaagaaaca gtaaacttta tgattcttaa aaaaaa      4436
```

```
&lt;210&gt; 22
&lt;211&gt; 2040
&lt;212&gt; DNA
&lt;213&gt; Homo sapiens

&lt;220&gt;
&lt;221&gt; misc_feature
&lt;223&gt; Incyte ID No: 7131221CB1

&lt;400&gt; 22
cacagagtga acaagagaga gtcatttggg aaacaaaagg agaattttac agagagagag 60
ggatagctaa aactacgtga gcctggcgag ggtgcagagc agaaagtaga gactgtccga 120
agactgctat ctgggacgag acaagttgtt aaagggacag gagagaaagc agagctattt 180
caagagtgag ccacagaagg gaatccagag gccatctaag cgaggaaggg tctacaggca 240
gtgagtgaag gccaggagca gggcccaggc caggcacgac caccgagggg atgaacttca 300
cagtgggttt caagccgctg ctaggggatg cacacagcat ggacaacctg gagaagcagc 360
tcatctgccc catctgcctg gagatgttct ccaaaccagt ggtgatcctg ccctgccaac 420
acaacctgtg ccgcaaatgt gccaacgacg tcttccaggc ctcgaatcct ctatggcagt 480
cccggggctc caccactgtg tcttcaggag gccgtttccg ctgcccatcg tgcaggcatg 540
aggttgtcct ggacagacac ggtgtctacg gcctgcagcg aaacctgcta gtggagaaca 600
ttatcgacat ttacaagcag gagtcatcca ggccgctgca ctccaaggct gagcagcacc 660
tcatgtgcga ggagcatgaa gaagagaaga tcaatattta ctgcctgagc tgtgaggtgc 720
ccacctgctc tctctgcaag gtcttcggtg cccacaagga ctgtgaggtg gccccactgc 780
ccaccattta caaacgccag aaggacaata gccggaggca gaagcagttg ttaaaccaga 840
ggtttgagag cctgtgcgca gtgctggagg agcgcaaggg tgagctgctg caggcgctgg 900
cccgggagca agaggagaag ctgcagcgcg tccgcggcct catccgtcag tatggcgacc 960
acctggaggc ctcctctaag ctggtggagt ctgccatcca gtccatggaa gagccacaaa 1020
tggcgctgta tctccagcag gccaaggagc tgatcaataa ggtcggggcc atgtcgaagg 1080
tggagctggc agggcggccg gagccaggct atgagagcat ggagcaattc accgtaaggg 1140
tggagcacgt ggccgaaatg ctgcggacca tcgacttcca gccaggcgct tccggggagg 1200
```

aagaggaggt ggccccagac ggagaggagg gcagcgcggg gccggaggaa gagcggccgg 1260
atgggcctta aggcctgcgc cgacccgacc ctgctcgaga gcccgcgcta gagtcgggga 1320
ggatctgcgc agagaccgca gcatcaccca aatcggcgcc ggccccggga ggatctcaat 1380
aaagaactcg agcgtcccag acccgtatct cctttcgctg cccaaccccg cagcctgggc 1440
ttcgaaggcg acccgcccac catcctgccc ttcccagaac ctgagaccgt ctgggggcg 1500
gaagccaaat gaacccctat tgggcacctc tgtgatgcca ggagcgaact ggtgagccca 1560
gcgccctggg aagagggccg agggcggggc ggtggtgccg ggacctctga ggtcctgggg 1620
atttggggac ccttggggtc cacatgcacc tggctgacct ggctgaaagc cgctgtctcg 1680
gagcccccca cagcattttg ttcccctccc gctggcccgg gggccccacc ttcccacggg 1740
ttcccacgct gctgtgactg ccctgcctct acgacaaaag ccaacgggtc ttcagtactt 1800
ttattaaaaa atagtcacgc agacagtgcc ctggtggctc tgccccgcat cccaactctg 1860
gggtggggga aaggggtcaa cgttttcgca gccccaaacc gggccatcac ttgcccaccg 1920
agtcgaatat gatgcggttc tgctcggcgc gctcccgctg gctctgcgtc cgcgccagct 1980
ccagcagggt ccgcagcagg tgaaaggtga ggtcaatgga cagagaaggg ttgtccgcgc 2040

<210> 23
<211> 2067
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7480551CB1

<400> 23
ggagctggga gggagcttta agggcggac gggcgggagg tcggggtcct ccggggatta 60
gagccggtgg gctcgttgtg ggcgccattt ctcggcgtct cccgaggagc cgcccctttc 120
tcagccttgc tcggctcttc cccgctctgg tcgccggggc tgcgccgtcc ccagctcagt 180
gacaaaaatg ctgagtttct tccgtagaac actagggcgt cggtctatgc gtaaacatgc 240
agagaaggaa cgactccgag aagcacaacg cgccgccaca catattcctg cagctggaga 300
ttctaagtcc atcatcacgt gtcgggtgtc ccttctggat ggtactgatg ttagtgtgga 360
cttgccaaaa aaagccaaag gacaagagtt gtttgatcag attatgtacc acctggacct 420
gattgaaagc gactattttg gtctgagatt tatggattca gcacaagtag cacattggtt 480
ggatggtaca aaaagcatca aaaagcaagt aaaaattggt tcaccctatt gtctgcatct 540
tcgagttaag ttttattcct cagaaccaaa taaccttcgt gaggagctaa cccggtattt 600
atttgttctt cagttaaaac aagatattct cagtggaaaa ttagactgtc cctttgatac 660
agcagtgcaa ttggcagctt ataatctgca agctgaactt ggtgactatg atcttgctga 720
gcatagtcct gaacttgtct cagagttcag attcgtgcct attcagactg aagagatgga 780
actggctatt tttgagaaat ggaaggaata cagaggtcaa acaccagcac aggctgaaac 840
caattatctg aataaagcca aatggctaga aatgtatggg gttgatatgc atgtggtcaa 900
ggctagagat gggaatgact atagtttggg actaacacca acaggagtcc ttgtttttga 960
aggagatacc aaaattggct tatttttttg gccgaagata accagattgg attttaagaa 1020
gaataaatta accttggtgg ttgtagaaga tgatgatcag ggcaaagaac aggaacatac 1080
atttgtcttt agactggatc atccaaaagc atgcaaacat ttatggaaat gtgctgtgga 1140
gcatcatgct ttcttccgcc ttcgaggccc cgtccaaaag agttctcatc gatcaggatt 1200
tattcgacta ggatcacgat ttagatatag tgggaaaaca gagtatcaga ccacaaaaac 1260
caataaagca agaagatcaa catcctttga aagaaggccc agcaaacgat attctagacg 1320
aactctacaa atgaaagcat gtgctacaaa acctgaagaa cttagtgttc acaataatgt 1380
ttcgacccaa agtaatggct cccaacaggc ttgggggatg agatctgctc tgcctgtgag 1440
tccttccatt tcctctgctc ctgtgccagt ggagatagag aatcttccac agagtcctgg 1500
aacagaccag catgacagga aatggctctc tgctgccagc gactgctgtc aacgtggtgg 1560
aaaccagtgg aacacaaggg ccttgccccc accccagacc gcacatagaa actacactga 1620
ctttgttcat gagcacaatg tgaagaatgc aggaatccgt catgatgttc attttcctgg 1680
ccatacagcc atgactgaga tatgagtgtt gagcctctta ggctttggga ctctttgtca 1740
tgcaagttga tggtatacat tatctggtgt ttataaagga ttaatcacat taggagtatt 1800
tgggagaatt tacagtgagt cactagttgt tcagtgctgt ttgtaattga attcttccat 1860

gaaagggaca aggaatcaag gaagccatat agcatcaatg ataatgacaa atgtttgtgt 1920
tgaaaagagt gtgtatacca ttgtggtttt ggaagagttt tcagacctta gtatgttcac 1980
acatcaccag actgtatctc aggagaaggt ttgtgtttgt gaacaaggtg cccattattc 2040
ccccaccaca tgccatccaa agagatc 2067

<210> 24
<211> 1640
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3315870CB1

<400> 24
gctgcaaacc ccactagcca gtgtcagcct ctcggcggga ggaggcggcg gcggaggagg 60
agcaggggga gggctgtcaa attcgggagc cagatttttt cccttctcct ggcaatccct 120
tccgcttccc cggctcccgg cgtgacatct gcgggccggg gacctgcatg tgtgtgcgcg 180
cgaaggagcg gaagaatggc agtgctcaaa ctcaccgacc agccaccatt ggttcaggca 240
atcttcagcg gtgatccaga ggagatccgg atgctcatcc ataaaactga agatgtgaat 300
actctggatt ctgagaaacg aacccctctt catgtggccg catttctggg agatgcagag 360
atcattgaac tcctgatttt gtcaggagct cgtgtaaatg ccaaggacaa catgtggctg 420
actccactgc accgggctgt tgcttccaga agtgaagaag cagtacaggt tttgattaag 480
cactcagctg atgtcaatgc aagggacaag aactggcaga cccctcttca tgtggcagca 540
gccaacaagg ctgtcaaatg tgcagaagtg atcattcccc tgctgagcag tgtcaatgtc 600
tccgaccgag gggggcgcac agccttgcac catgcggctc tgaacggcca cgtggagatg 660
gtcaatttac tcttggccaa aggggcaaat atcaatgcat ttgacaagaa ggaccggcgt 720
gctctgcact gggcagcata catgggccac ttggatgttg tagcattgct cattaaccat 780
ggcgcagaag tgacctgtaa ggataagaag ggttataccc ctctgcatgc tgcagcctcc 840
aatggacaga ttaatgttgt caagcatctc ctgaacctgg gggtggagat tgatgaaatc 900
aatgtctatg gaaatacagc gcttcacatc gcctgctaca atggacagga tgctgtggtt 960
aacgagttga ttgactacgg tgctaacgtg aaccagccaa acaataatgg gttcacccct 1020
ttgcattttg ctgctgcctc cactcatggt gctttgtgtc ttgaattgtt agtaaacaac 1080
ggggcagatg ttaacattca gagtaaagat ggcaaaagtc cactgcacat gacagctgtc 1140
catggaaggt tcacacggtc acagaccctc attcagaatg gaggtgaaat tgactgtgtg 1200
gataaggacg gcaacactcc tctccatgtg gctgcaagat acggtcatga gcttttgatt 1260
aacaccttaa taaccagcgg agctgacaca gccaagtagg ttaccgcaaa aatacggtgg 1320
aataattgcc tcaagtggga atactgccaa aaagattctt ccgtgcagta gatagttccc 1380
atttatccag gttaaggtgg atctatacca ttatactaaa tacaaattaa attttaaata 1440
aattaagtgc cttttaatga cagaggcaaa gaagaaccaa ttttattttt tagcttcatc 1500
caaatgaggt ctatttcagt ggttttaatt aaggaaactt gaactttatt cgtaactttt 1560
ctttctaata ttcttctgtc cttcccaatg cttcatatta aacaggaaaa ataaaaccta 1620
actgagccaa tttagaatgt 1640

<210> 25
<211> 1497
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7484690CB1

<400> 25
atgagggaaa tcgtgctcac gcagaccggg cagtgcggga accagatcgg ggccaagcag 60
ttctgggagg tgatctctga tgaacatgcc atcgactccg ctggcaccta ccacggggac 120

agccacctgc cgctggagcg cgtcaacgtg caccaccacg aggccagcgg tggcaggtac 180
gtgcctcgcg ctgtgctcgt ggatctggag ccgggcacca tggactccgt gcgctcgggg 240
cccttcgggc aggtcttcag gccagacaac ttcatttccc gtcagtgtgg ggccggaaac 300
aactgggcca agggacgcta caccgaaggc gcggagctga cggagtcagt gatggacgtt 360
gtcagaaagg aggctgagag ctgtgactgc ctgcagggtt tccagctgac ccactccctg 420
ggtgggggga ctgggtctgg gatgggtacc cttctgctca gtaagatccg ggaggagtac 480
ccagacagga tcataaacac attcagcatc ctgccctcgc ccaaggtgtc ggacaccgtg 540
gtggagccct acaacgtcac cctctcagtc caccagctca tagaaaacgc ggatgagacc 600
ttctgcatag ataacgaagc gctatatgac atatgttcca ggaccctaaa actgcccaca 660
cccacctatg gtgacctgaa ccacctggtg tctgctacca tgagtggggt caccacgtgc 720
ctgcgcttcc cgggccagct gaatgctgac ctgcggaagc tggccgtgaa catggtcccg 780
tttccccggc tgcatttctt catgcccggc tttgccccac tgaccagccg ggcagccag 840
cagtaccggg ccttgactgt ggctgagctt acccagcaga tgtttgatgc taagaacatg 900
atggctgccc gtgaccctg tcacggccgc tacctaacgg tggctgccat tttcaggggt 960
cgcatgccca tgagggaggt ggatgaacag atgttcaaca ttcaagataa gaacagcagc 1020
tactttgctg actggttccc cgacaacgta aaaacagccg tctgtgacat cccacccgg 1080
gggctaaaaa tgtcagccac cttcattggg aacaacacag ccgtccagga actcaagcgg 1140
gtctcagagc agtttacagc aacgttcagg cgcaaggcct tcctccactg gtacacgggc 1200
gagggcatgg atgagatgga attcactgag gccgagagca acatgaacga cttggtgtct 1260
gaatatcagc aatatcagga tgccacggcc gagggaggag gagtatgagg aggaggaggt 1320
ggcctagaac tctccttttc taggtaaagg ggggaagcag tgtggatcct tcactgtgtt 1380
ctgacagcca tgtgtcacta tgcgctcgtt catttgtgtc ttcacatctc ctgctgcatt 1440
ttaaagcatt tttatagtat gcggttttgc ctaataaagt attctcacag cgaaaaa 1497

<210> 26
<211> 2065
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7612559CB1

<400> 26
ccgagatccg cgctctctac aacgtgctgg ccaaagtgaa gcgggagcgg gacgagtaca 60
agcggaggtg ggaagaggag tacacggtgc ggatccagct gcaagaccgt gtaaatgagc 120
tccaggagga agcccaggag gctgatgcct gccaggagga gctggcactg aaggtggaac 180
agttgaaggc tgagctggtg gtcttcaagg ggctcatgag taacaacctg tcggagctgg 240
acaccaagat ccaggagaaa gccatgaagg tggatatgga catctgccgc cgcatcgaca 300
tcaccgccaa gctctgcgat gtggctcagc agcgcaactg cgaggacatg atccagatgt 360
tccaggtccc atccatgggg gggcggaagc gggagcgcaa ggctgccgtc gaggaggaca 420
cctccctgtc ggagagtgag ggccccgcc agcccgatgg ggatgaggag gagagcacag 480
ccctcagcat caacgaggag atgcagcgca tgctcaacca gctgagggag tatgattttg 540
aggacgactg tgacagcctg acttgggagg agactgagga gaccctgctg ctttgggagg 600
atttctcagg ctatgccatg gcagctgcag aggcccaggg agagcagcag gaagatagcc 660
tggagaaggt gattaaagat acggagtccc tgttcaaaac ccgggagaag gagtatcagg 720
agaccattga ccagatagag ctggagttgg ccacggccaa gaacgacatg aaccggcacc 780
tgcacgagta catggagatg tgcagcatga agcgcggcct ggacgtgcag atggagacct 840
gccgccggct catcacccag tctggagacc gaaagtctcc tgctttcact gcggtcccgc 900
ttagcgaccg ccgccgccgc caagcgaggc tgaggactcc gatcgcgatg tctcatctga 960
cagctccatg agatagagac ctgcctcccc cttgcacccg aggccctcgc agcagggagc 1020
tcagcgaggc agagggtggg gctgcacaga ggggaacatc agctgcagct ctgcaccagg 1080
ccggtccctg ggactgggg cgctcctccc tcaggctttc tccctcagtc ttggcttctc 1140
cagggctctg ggtgtctgg agctaggctt ggccctacca ttctggggcc atttccacca 1200
cagttggggc tctcctgcct tcacgcgtgg gtgtctgcta cttccccatc tttaaaatgc 1260
tgccagagcg attgcggccc ctcaccttgt ccacgtatca ggaatgtgaa tgtgggacct 1320

ttcctccatc cctgttgtcc ggagccagct cactgtcttc cacactggtg ctaactggcc 1380
caggcactgg agtggaatag aatgcagctg gaggctacgc atggcctctg cagcacacgc 1440
agctggagag ggcttctgtc cctgtcagcg gcagagggcg ttggggctgg ccggggcacc 1500
ttgtccctgc tatggtccac atgctcacgc tgtccacctg ccaggtggag tgtatgtggc 1560
tgtggccctc cctcgtggag gtgccgtgct ttaaagaggc cttagtgccc gggatgggca 1620
cagtgttttg aagggaggtg ggagctcttg ctctcctggt cactgcagaa tgacagagaa 1680
ggtgaagctc catgcatgtg tgcgcgggtg tatgtgcgct cagggtctct gtttaagtat 1740
cagctaaaga tgtgcttcct ccgtgtctgt catacactga gaccaacagg ctacagtgtc 1800
cctgattctt ggaaaagcct ggagaagctg gggagatgcg gttcacaatg cctcggtata 1860
ggaggctgtg ttgagctgac attcaaatgg attctttaat aataatgaaa ctggcgagta 1920
tttattgtgc actttggtgt ccctgtctcc agcacttcct aatattcact agtttgaact 1980
ctgaggtagg tacttttttt ttttgagatg gagtctcata ctctgttgcc taggctggag 2040
tgcagtggtg cgatcacagc tcgct 2065

<210> 27
<211> 762
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4940751CB1

<400> 27
gcagaggcag catagcagca gccagctcca tccatcctct ttcccctcct cgcttcgctt 60
cctcggcgga ttcctcctcc ctcgacagtc cccgtcgccg tccccttccg gtgcgcaagt 120
cgcccgagat ggcaaacgcg agatcgggtg tcgctgtgaa tgacgagtgc atgctcaagt 180
tcggcgagct gcagtcgaag aggctgcacc gcttcctaac tttcaagatg gacgacaagt 240
tcaaggagat cgttgtggac caggtcgggg atcgcgctac cagctacgag gacttcacaa 300
acagcctccc cgagaatgac tgccgatacg cgatctatga tttcgacttt gtcactgcag 360
aagatgtcca gaagagcagg atcttctata tcctatggtc cccatcctcc gccaaggtga 420
agagcaagat gctttatgca agctcaaacc aaaaattcaa gagtgggctc aatggcattc 480
aggtggaact gcaggctact gatgcaagtg aaatcagcct tgatgagatc aaggatcggg 540
ctcgctaggc atcatcatga tcatgcatca tggacttggc ctactactgt ggatttgtat 600
gccattatag acttggtgct gtgaaagact gcttgatgat ttgcgggttt gttgctgtgt 660
aaaaaaaggt cccatggctc ccagaagacc atgaaggttc ggatctatca tgtaattcct 720
tgttatctgc gaattaatgt atagtgttgc attggtcgcg tc 762

<210> 28
<211> 2211
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7946761CB1

<400> 28
atgacctggg gcaccccgga ctttcttaat cgtagctcca cccactcgag ccgggtgcct 60
tcgcgtttcc cgtttttaaa tgagatagtg gcacacccgg tggcatcctc ccacccgggc 120
tcttatcggc ggtcccagac cctgcttgag cgcctccggg tgtcaagggc ccctgaggac 180
actaaagctc tcgaaccccg atgtggaccc ccgtgcggcg cggggcagcc tggctgggaa 240
ccctgctcgg ccctggagag gggccccccg agccgagggg aggagcggcg catgcccaca 300
agccccccgg cgggaagtag gaaatcgacc gaccaggcgg tgcgcttcgg acccagccag 360
ggcatgtgct cggaggcccg cctggctcgc aggttgcggg atgcgctgcg ggaggaggag 420
ccgtgggcag tagaggagct gctgcgctgc ggcgcggacc ctaatttggt gctagaggac 480

ggcgcagcgg ctgtgcactt ggcggccgga gcccggcacc cgcgcggcct gcgttgcctc 540
ggggccctac tgcgccaagg cggggacccc aacgctcgat ctgtcgaggc actgacgccg 600
ctgcatgtgg ccgccgcgtg gggctgccgc cgcggcctgg agctgctgct gagccaagga 660
gcggacccgg cgctgcgcga ccaggacgga ctccggccgc tggacctggc cctgcagcag 720
ggacacctgg agtgcgcgcg agtcctgcag gatctcgaca cgcggaccag gacccggacc 780
cggatcgggg cagagactca ggagcccgag cctgcacctg gcaccccagg cctctctgga 840
cctaccgatg agacgctgga ctccatagca ctccaaaagc agccatgcag aggtgacaac 900
agggacattg gcttggaggc tgacccagga ccccccagcc tccctgttcc ccttgaaact 960
gtggacaaac atgggagctc ggcgtcccct ccagggcact gggattacag ctcagacgcc 1020
tctttcgtca cagcggttga ggtctctgga gctgaggacc cagcctcgga cactcccccc 1080
tgggctgggt cattgccacc gaccaggcag ggacttctgc atgttgtcca tgccaaccag 1140
agggtaccta ggtctcaggg cacggaggca gaactgaatg cccgtctgca ggccctgact 1200
ctgaccccac caaatgctgc tggcttccag tcctcccctt cctccatgcc tctcctggac 1260
aggagtccag ctcatagccc cccacggaca ccaacccctg gagcttctga ctgccactgc 1320
ctgtgggagc accagacatc cattgatagt gacatggcca cgctctggct gacagaggat 1380
gaggcaagct ctacaggtgg cagggaacct gtcggccctt gccggcacct gccagtctcc 1440
actgtgtctg acttggagtt gctgaaggga ctccgagcac ttggtgagaa tcctcacccc 1500
atcacaccct tcaccaggca gttgtaccac cagcagctgg aagaagccca gattgctcca 1560
ggcccagagt tttcagggca cagcctagaa ctggctgcag ccctgcggac gggctgtatt 1620
ccagatgtcc aggcagatga agacgcgctg gcccagcagt ttgagcggcc agatcctgcc 1680
aggaggtggc gggagggggt cgtgaagtct agcttcacgt atctgctgct ggaccccagg 1740
gagactcagg acctgccagc ccgagccttc tcactgaccc cagctgagcg ccttcagact 1800
ttcatccgtg ccatcttcta cgtgggcaaa gggacgaggg cccggccata tgtccacctc 1860
tgggaggccc ttggtcacca tgggcggtca agaaaacagc cccaccaggc ctgccccaag 1920
gtgcgtcaga tcttggacat ctgggccagt ggttgcggcg ttgtgtccct acattgcttc 1980
cagcacgtgg tcgctgtgga ggcttataca cgggaggcgt gtattgtgga agccctaggg 2040
atccagacgc tcaccaacca gaagcaaggg cactgctatg gagtggtggc aggttggcca 2100
cctgctcgtc gccggcgctt ggggtgcac ctgctgcacc gtgccctcct tgtcttcctg 2160
gctgaaggcg agcgacagct tcatccccag gacatccagg cccggggctg a 2211

<210> 29
<211> 1634
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3288747CB1

<400> 29
ctcagctaag ggtcagcatc ttatccccac tttctggcct ccccaccatg agccgccaat 60
tcacctacaa gtcgggagct gctgccaagg ggggcttcag cggctgctcc gctgtgctct 120
caggggggcag ctcatcctcc taccgagcag gggggcaaagg gctcagtgga ggcttcagca 180
gtcggagcct ttacagcctg gggggtgccc ggagcatctc tttcaatgtg gccagtggca 240
gtgggtgggc aggaggctat ggatttggcc ggggccgggc cagtggcttt gctggcagca 300
tgtttggcag tgtggccttg gggtccgtgt gtccgtcgtt gtgcccgccc gggggtatcc 360
atcaggtcac catcaacaag agcctcctgg cacccctgaa cgtggagctg gaccctgaaa 420
tccagaaagt gcgtgcccag gagcgggagc agatcaaggt gctgaacaac aagttcgcct 480
ccttcattga caaggtgcgg ttcctggagc agcagaacca ggtgctggag accaagtggg 540
agctgctaca gcagctggac ctgaacaact gcaagaataa cctggagccc atccttgagg 600
gctacatcag caacctgcgg aagcagctgg agacgctgtc tggggacagg gtgaggctgg 660
actcggagct gaggagcgtg cgcgaagtgg tggaggacta caagaagaga tacgaagaag 720
aaataaacaa gcgcacaact gctgagaatg aatttgtggt gcttaagaag gacgtggacg 780
cagcttacac gagcaaagtg gagctgcagg ccaaggtgga tgccctggat ggagaaatca 840
agttcttcaa gtgtctgtac gaggggggaga ctgctcagat ccagtcccac atcagcgaca 900
cgtccatcat cctgtccatg gacaacaacc ggaacctgga cctggacagc atcattgctg 960

aggtccgtgc ccagtatgag gagatcgccc ggaagagcaa ggccgaggcc gaggccctgt 1020
accagaccaa gttccaggag ctgcagctag cagccggccg gcatggggat gacctgaaac 1080
acaccaaaaa tgagatctca gagctgaccc gtctcatcca aagactgcgc tcggagattg 1140
agagtgtgaa gaagcagtgt gccaacctgg agacggccat cgctgacgcc gagcagcggg 1200
gggactgtgc cctcaaggat gccagggcca agctggatga gctggagggc gccctgcagc 1260
aggccaagga ggagctggca cggatgctgc gcgagtacca agagcttttg agcgtgaagc 1320
tgtccctgga tattgagatc gccacctacc gcaagctgct ggagggcgag gagtgcagga 1380
tgtccggaga atataccaac tccgtgagca tttcggtcat caacagctcc atggccggga 1440
tggcaggcac aggggctggc tttggattca gcaatgctgg cacctacggc tactggccca 1500
gctctgtcag cgggggctac agcatgctgc ctgggggctg tgtcactggc agtgggaact 1560
gtagcccccc agtggtcagc aatgtcacca gcacaagtgg cagctctggc agtagccgtg 1620
gagtttttgg aggg 1634

<210> 30
<211> 4706
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 8200016CB1

<400> 30
catgtgaaca ccaattagag ctgactattc ccgggattgt ggtactcggg gctgtgtcaa 60
tcaagggtgc tacaatagca cgtgcaccag tggtgcctca agacccaccg gggagaggct 120
tatcttaact ccagctgccg aatgagaatg agtttgaagc tttttgcagg atcatggaac 180
agagcctcca tgcaatagtg catcctgagg taaactgtta cctgagtaag ggctttaagt 240
aatgcatttc ctgggaacga cagttgtgac agaagagaat gctggaaccc gtagcaagat 300
tcctgtctga gatggaaaga tgtctcacta tcattttatc aagtgctgtt gctttcagct 360
atgtaacgtt tttcgatccc atgagatgga aatcgaccag tgcttgctag agtcccttcc 420
ccttggccaa cggcagcgtc tagtgaagcg catgcgctgt gagcaaatca aagcctacta 480
tgagcgcgag aaggcttttc agaagcagga agggttcctg aaaaggctga agcatgcgaa 540
gaatccgaaa gttcacttca acctcacgga catgctacag gacgcgatta tccaccacaa 600
tgacaaagaa gtgcttcggc tcctgaagga gggggcagac ccccacaccc tcgtctcctc 660
gggagggtcc ctgctccatc tgtgtgctcg gtatgataat gccttcattg cagaaattct 720
gattgacaga ggagtcaacg tcaaccacca ggatgaagac ttctggacgc ccatgcacat 780
tgcctgtgcc tgcgataacc ctgatattgt cctgcttctt gtattagctg gagccaatgt 840
ccttctccag gatgtgaatg gaaatatccc attagattat gctgtagaag ggacagaatc 900
cagctctatc ctgttgacct atctggatga aaatggagtg gatttgacct cactgcgcca 960
gatgaagctt cagagaccaa tgagtatgtt aacagatgtc aaacacttct tatcatctgg 1020
aggaaatgtc aatgagaaaa acgatgaagg agtaaccctg ttacacatgg cgtgtgcgag 1080
tggctacaag gaggtggtgt ctcttatcct ggaacatggt ggagacctca acatagtaga 1140
tgatcagtac tggactcccc tccacttggc agccaaatat ggccagacaa atctggtgaa 1200
acttctcctg atgcatcagg caaacccaca cctcgtgaac tgtaatgagg agaaggcgtc 1260
agatattgct gcctctgagt ttattgagga aatgctgctg aaagccgaaa ttgcctggga 1320
agaaaaaatg aaagagcctt tatctgcttc taccttagct caagaagagc cctatgaaga 1380
gatcattcac gatcttcccg tactgtcgag taagctcagt cccctggtgt taccaattgc 1440
caagcaagac agtttgttgg aaaaagacat tatgttcaaa gatgcaacaa aaggtctgtg 1500
taagcagcag tctcaggaca gcatccctga aaaccccatg atgagcggtt ccaccaaacc 1560
cgagcaggtc aagctaatgc ctcctgcccc aaacgatgac ctggcaacgc tcagcgagct 1620
caatgatggc agcctgctct atgagattca gaagcgcttt gggaacaatc agatctatac 1680
attcattgga gacattcttt tgcttgttaa cccatacaag gagcttccaa tttattcttc 1740
catggtgtcc cagctgtatt tcagctcctc agggaagctg tgttcctcgc tgcctcctca 1800
cctcttctcc tgtgtggaga gagcctttca ccagctcttc cgggaacagc ggcctcagtg 1860
tttcatcctc agtggagaaa ggggatcagg aaagtctgaa gccagcaaac aaatcataag 1920
acacctcacc tgcagggctg gcgccagcag ggccacactg gattccagat tcaaacatgt 1980

cgtgtgcatc ttagaagcct ttggacatgc caagaccaca cttaatgatt tgtccagttg 2040
cttcatcaag tattttgaac tgcagttctg tgagaggaaa caacagctaa ccggagccag 2100
aatttataca tatttgctag agaaatccag acttgtttca caacctcttg gccagagcaa 2160
ttttctcatt ttctacttgt tgatggatgg gttatctgct gaagaaaaat atggacttca 2220
tcttaataat ttatgtgcac accggtattt gaaccagacc atacaggatg atgcatccac 2280
aggggagcgt tctctgaaca gggagaaatt ggctgttttg aaacgagccc tgaatgtagt 2340
tggcttcagc agcttggagg tggagaatct gttcgtaatt ctagcagcaa tattgcacct 2400
tggagacatt cggtttactg ccctgaatga ggggaactcc gccttcgttt ctgacctcca 2460
gctcctggaa caagtggctg gaatgttaca agtatcaaca gatgaattgg catctgcctt 2520
aacaactgat attcaatatt ttaaagggga tatgataata cgacgacata ccatacagat 2580
tgctgagttt ttccgagacc tcttggccaa gtccctgtac agtcgtttgt ttagcttttt 2640
ggtgaatacc atgaattctt gcctccacag tcaagatgaa cagaaaagca tgcagacatt 2700
ggatattgga atattggaca tttttggttt tgaagagttt caaaagaatg aatttgaaca 2760
actttgtgtc aacatgacca atgagaagat gcaccactat atcaatgaag tgctttttct 2820
ccacgagcaa gtggaatgtg tacaagaggg agttaccatg gaaacagcat attctgctgg 2880
taaccagaat ggagttttgg actttttttt ccagaagcca tctggatttc tcaccttatt 2940
ggatgaagaa agtcaaatga tttggtcagt ggaatcaaat tttccaaaaa aactacaaag 3000
tctcctagaa tcctcaaaca caaatgcggt gtactccccc atgaaggatg ggaatgggaa 3060
tgttgccctc aaagaccacg gtacagcctt caccatcatg cactacgcag gaagggtaat 3120
gtatgatgtt gttggggcga ttgaaaaaaa taaagactcc ctttcacaga atcttctatt 3180
tgtaatgaaa actagtgaaa atgtcgtgat caatcatttg ttccagtcga aattgtcaca 3240
aacaggatcc ctcgtatctg cctatccttc ctttaaattc cgaggacata agtctgccct 3300
gctcagtaag aaaatgacag cttcttcaat tattggagaa aacaagaatt atctagaact 3360
tagtaagtta ttaaaaaaga aaggaacttc tacatttctt caaagattgg aacgaggaga 3420
tccagtcacc atagcatcac aactcaggaa atcactaatg gatattattg gaaaacttca 3480
gaagtgcact ccacacttca ttcattgcat caggcccaat aactcaaagc tgccagatac 3540
ttttgataat ttttacgtgt ctgctcagct acaatatatt ggggtcctgg agatggtgaa 3600
gatcttccga tatggatacc ctgttcgcct ttccttctcg gatttcctgt caaggtataa 3660
gccactggct gatacattcc tgcgtgagaa gaaggaacag tcagctgccg agcgatgtcg 3720
acttgttctc cagcagtgta aattacaagg ctggcagatg ggagtccgaa aagtgtttct 3780
aaaatactgg catgctgacc aactcaatga tttgtgccta cagttgcaga gaaaaattat 3840
aacctgccaa aaagttatca gaggattttt agcacgccag cacctgcttc agagaatgag 3900
catcagacaa caagaggtga cttctatcaa tagctttctg cagaacacag aggacatggg 3960
gctgaaaacc tacgatgccc tggtcattca gaatgcttca gacattgccc gggaaaatga 4020
ccggctccgt agtgaaatga acgctcccta ccataaagag aagttagagg tcaggaacat 4080
gcaagaggaa ggaagcaaaa gaaccgatga caagagtgga cccaggcatt tccaccccag 4140
ctccatgtca gtctgcgcgg ccgtggatgg cctgggccag tgcctcgttg gcccgtccat 4200
ctggtctcct tcgctgcact cggtgttcag catggatgac agcagcagcc tcccgtctcc 4260
acggaaacag cccccgccca agccaaagag ggaccccaac acccggctga gtgcttccta 4320
tgaggctgtg agcgcctgcc tctccgcggc cagggaagcg gccaacgaag gtcagccttg 4380
gggagggacc cagcctcgtg ttccgggctc gcgcatgctc tgacttcgcc ttggggcgcc 4440
catggcagta ctgtcgccct aatgtattct taatagaaat aaatccaatt gttggcttgc 4500
cagcagctct taatcattaa atataaatat atttattcaa tctctaagcc tcttagggaa 4560
aagctactta catggcattt ccttaatccc atccccaacc tgctccaaga gcagtatcaa 4620
tcattcagaa agtcggagtt attcagttaa ggtcccatgg gaagttccca aaaaaaaacg 4680
gtcggctcct tccaccttta aattac 4706

<210> 31
<211> 3029
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3291962CB1

<400> 31
ctggctggag gttgacacag gagtgctcag gggagcagca tcacaagagg gcagatcgaa 60
agcatcgtcc ttgctgaaaa aatggcagag gtggaagcgg tacagctgaa ggaggaagga 120
aaccggcatt tccagctcca ggactacaag gccgccacaa atagctacag ccaggccctg 180
aagctgacca aggacaaggc cctgctggcc acgctttatc ggaaccgggc agcctgtggc 240
ctgaaaacgg agagctacgt ccaggcagct tcagatgcct ccagagccat cgacatcaac 300
tcctcggaca tcaaggctct gtatcggcga tgccaggcac tggagcacct ggggaagctg 360
gaccaggcct tcaaagacgt gcagcgttgt gccaccctcg agccacggaa ccagaacttc 420
caggagatgc tgaggagact caacaccagc attcaggaga aactccgagt gcagttctcc 480
acagactcga gggtacagaa gatgtttgag atcctcttgg atgaaaacag tgaggctgat 540
aagcgggaaa aggctgccaa caatctcatt gtcctaggcc gtgaggaagc aggggctgag 600
aagatcttcc agaacaatgg agtagccttg ctactgcagc ttctggacac taagaagcct 660
gagctggtgc tggctgcagt gcggaccctg tcgggcatgt gcagcggcca ccaagccaga 720
gccacagtga ttctgcatgc agtgcggata gaccgaatct gtagcctcat ggccgtggag 780
aatgaggaga tgtctctggc tgtctgcaac ctgctccaag ccatcattga ctccttgtct 840
ggggaggaca agcgggagca tcgagggaag gaggaggccc tggttctaga caccaagaag 900
gacctgaagc agatcaccag ccacctgctg gacatgctag tcagcaagaa ggtgtctggc 960
cagggcaggg atcaggcgct gaacctgctc aataagaatg ttcccaggaa ggaccttgcc 1020
attcatgaca actcacgtac catctatgtg gtggataatg gtctgaggaa gatcctgaag 1080
gttgtggggc aggttccaga tctgccatcc tgcctgcccc tgactgacaa cacccgcatg 1140
ctggcctcta tcctcatcaa caagctctat gatgacctgc gctgtgaccc ggagcgcgat 1200
cacttccgca agatctgtga ggaatatatc acgggcaagt ttgaccccca ggacatggac 1260
aagaacttga atgccatcca gacagtgtca gggatcctgc agggcccctt tgacctgggc 1320
aaccagctgc tgggactgaa aggtgtgatg gagatgatgg tggcactatg tggctcagag 1380
cgcgagacgg accagctggt ggccgtggag gccctcatcc atgcctccac gaagctcagc 1440
cgcgccacct tcatcatcac caatggagtg tcactgctca aacagatcta caagaccacc 1500
aaaaatgaga agatcaagat ccgcacactg gtgggactct gtaagctcgg ctctgcaggt 1560
ggcacagact acggtctcag gcagtttgcg gaagggtcga cagaaaaact ggccaaacag 1620
tgtcgcaagt ggctgtgcaa tatgtccata gacactcgga cccgacgctg ggcagtggag 1680
ggcctggcct acctcacgct ggacgctgat gtgaaggacg actttgtcca ggacgtccct 1740
gccctgcagg ccatgtttga gctggccaag accagtgaca agaccatcct gtactcggtg 1800
gccaccaccc tggtgaactg caccaacagc tacgatgtca aggaggtcat cccagagctt 1860
gtccagctcg ccaagttctc caagcagcat gtgcccgagg aacaccccaa ggacaagaag 1920
gactttatag acatgcgggt gaagcggctt ctgaaggcgg gtgtcatctc tgccctggct 1980
tgcatggtga aagcagatag tgccatcctc actgaccaga ccaaggagct gctggccagg 2040
gtattcctgg cactgtgtga caacccaaag gaccgaggca ccattgtggc tcaaggtggt 2100
ggcaaggccc tgattcccct ggctttggag ggcacagatg tgggcaaggt gaaggcagcc 2160
cacgctctag caaagatcgc tgctgtctcc aatccggaca ttgcttttcc tggggagcgg 2220
gtgtatgagg tggtgcggcc ccttgtaaga ctcttggaca cacagaggga tgggcttcag 2280
aactatgagg ctctcctagg cctcaccaac ctgtctgggc ggagtgacaa actccggcag 2340
aagatcttta aggagagggc cttgccagac atcgagaact acatgtttga gaatcatgat 2400
cagctgcggc aggcggccac cgagtgcatg tgcaacatgg tgctccacaa ggaggtacag 2460
gaaaggttct tggctgacgg gaatgaccgg ctgaagctgg tggtgctgct ctgcggggag 2520
gatgatgata aggtgcagaa tgcggctgca ggggctctgg ccatgctgac agcagcacac 2580
aagaaactgt gcctcaagat gactcaagtg acaacccagt ggttggagat cctccagcgg 2640
ctttgcctgc acgaccagct gtctgtccaa caccggggcc tggtcattgc ctacaaccta 2700
ctggcagccg atgctgagct ggccaagaag ctggtggaga gtgagctgct ggagatcctg 2760
actgtggtgg gcaaacagga gccagatgag aagaaggcag aagtggttca gacagcccga 2820
gaatgtctca tcaagtgcat ggattatggt ttcattaaac cagtgtctta gacagcgacc 2880
ctccgggatg ctgggagtgg tcctgtactg tgcagagtcc tgggttggtt gggttctcct 2940
gaagagtcag gtcatctagg gatcatagca gtgacaatga agtctcaata taaaggaaag 3000
acttgattgt tctctgaaaa aaaaaaaaa 3029

<210> 32
<211> 2074
<212> DNA

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1234259CB1

<400> 32
ctctcgcgag gacggacgcc attatcgcat ctccccgaca aacaccacga gaattccgca 60
gcccacacgg tgaccaaaag ccagccccac tgtgagttga actctttcgt gttgaccggc 120
cactctcctg tgctctggat gatgtcggaa cacgacctgg ccgatgtggt tcagattgca 180
gtggaagacc tgagccctga ccacccagtt gttttggaga atcatgtagt gacagatgaa 240
gacgaacctg ctttgaaacg ccagcgacta gaaatcaatt gccaggatcc atctataaag 300
tcattcctgt attccatcaa ccagacaatc tgcttgcggt tggatagcat tgaagccaaa 360
ttgcaagccc tggaggctac ttgtaaatcc ttagaagaaa agctggatct ggtcacgaac 420
aagcagcaca gccccatcca ggttcccatg gtggccggct cccctctcgg ggcaacccag 480
acgtgcaaca aagtgcgatg cgtcgtcccc cagactacag taatactcaa caatgatcgg 540
cagaacgcca ttgtagccaa gatggaagac cccttgagca acagggcacc ggattccctg 600
gaaaatgtca ttagcaacgc tgtgcctggg cgtcggcaga acaccattgt ggtgaaggtg 660
ccgggccaag aagacagcca ccacgaggac ggggagagcg gctcggaggc cagcgactct 720
gtgtccagct gtgggcaggc gggcagtcag agcatcggga gcaacgtcac gctcatcacc 780
ctgaactcgg aagaggacta ccccaatggc acctggctgg gcgacgagaa caaccccgag 840
atgcgggtac gctgcgccat catcccctcc gacatgctgc acatcagcac caactgccgc 900
acggccgaga agatggcgct aacgctgctg gactacctct tccaccgcga ggtgcaggct 960
gtgtccaacc tctcggggca gggcaagcac gggaagaagc agctggaccc gctcaccatc 1020
tacggcatcc ggtgtcacct tttctataaa tttggcatca cagaatccga ctggtaccga 1080
atcaagcaga gcatcgactc caagtgccgc acggcgtggc ggcgcaagca gcggggccag 1140
agcctggcgg tcaagagctt ctcgcggaga acgcccaact cgtcctccta ctgcccttca 1200
gagccgatga tgagcacccc acctcctgcc agcgagctcc cgcagccaca gccgcagccg 1260
caggccctgc actacgcgct ggccaacgca cagcaggtgc agatccacca gatcggagaa 1320
gacggacagg tgcaagtaat cccacaggga cacctccaca tcgcccaggt gccgcagggg 1380
gagcaagtcc agatcacgca ggacagcgag ggcaacctcc agatccatca cgtggggcag 1440
gacggtcagc ttctagaggc cacccgcatc ccctgcctcc tggccccatc cgtcttcaaa 1500
gccagcagtg gccaggtgct gcaggtgca cagctgatcg ccgtggcctc ctcggacccc 1560
gcggcagcgg gcgtggatgg gtcgccactc cagggcagcg acatccaggt tcagtacgtg 1620
cagctggcgc cagtgagtga ccacacggcc ggggcacaga cggccgaagc cctgcagccc 1680
acgctacagc cggagatgca gctcgagcac ggggccatcc agattcagtg agcggtgccc 1740
atggcaccag gagcccctcg ccggctccgc ctacggcccg gcccccacgc gccctgctct 1800
cacggcctcg gcacaggcag cggctgcacg tgttctgctg aagtgcgtct gaaggccgct 1860
gcctccgcgg ggaacagcat cctatcaact gaaagagcag ccgccgccgc ccccagccgg 1920
agaccccttt cgtttgagtc ctgctgttgg tgtcggagca cgaggggagg cacggtgcgg 1980
agagcgtcgc atatgcgcgg gaaatcaaga actatgatat ttttctgttt aaacagcttt 2040
ttttaatttg ctatggtgtt tataacaaaa aaaa 2074

<210> 33
<211> 2710
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1440608CB1

<400> 33
atggccaagt ttgccctgaa tcagaacctg cccgacctgg gcggcccccg cctgtgcccg 60
gtccccgccg ccgggggcgc acgcagcccg agctcgccct actcggtgga gacgccctac 120
ggcttccacc tggacctgga cttcctcaag tacatagagg agctggagcg tggccccgct 180

gcccgccgcg ccccgggacc cccgacctcg cgccgtcccc gcgcgccccg gcccggcctc 240
gcgggcgcac gtagcccagg cgcctggaca tccagcgagt ccctggccag tgacgacggt 300
ggagcaccgg gcatactctc ccagggcgcg ccctcggggc tcctgatgca gccgctgtcg 360
ccgcgcgcgc ccgtgcgcaa cccgcgcgtc gagcacacgc tccgggagac cagccggcgg 420
ctggagctgg cgcagacaca cgagcgcgcg cccagccccg gccgcggggt cccgcgcagc 480
ccacgcgggt ccggccgcag cagccccgcc cctaaccttg cccctgcttc gcccggccct 540
gcccaactgc agctggtgcg cgagcagatg gccgcggcgc tgcggcgcct gcgcgagctc 600
gaggaccagg cgcgaacgct gcccgagctg caggagcagg tgcgcgcgct gcgcgccgag 660
aaggcgcggc tgctggccgg gcgcgcgcag cccgagccgg acggggaggc tgagacgcgc 720
ccggacaagc tcgcccagct gcggcggctc accgagcgcc tggccacctc cgagcgcggc 780
ggccgtgcca gggccagccc ccgggctgac agcccagacg gcctggctgc agggcgcagc 840
gagggcgcgc tccaggtcct cgacggggag gtcgggagtc tcgatgggac gccccagacc 900
cgggaggtgg ccgccgaggc cgtgccccgag acccgagaag cgggtgccca ggccgtgccg 960
gagacccggg aggccggcgt ggaggctgcc cccgagaccg tggaggcgga cgcgtgggtg 1020
accgaggcgc tgctggggct gcctgcggcc gccgagcgcg agctagagct gctgcgcgcc 1080
agtctggagc accagcgcgg ggtgagtgag cttctgcggg gccggttgcg ggagctggag 1140
gaagcccgcg aggctgcgga ggaggcagcg gcgggggccc gggcccagct acgcgaggcc 1200
accacccaga ccccgtggag ctgtgccgaa aaggccgcgc agaccgagtc cccggcagag 1260
gcgccctcct tgactcagga gagctcgccc ggatccatgg acggagacag ggccgtggcg 1320
cccgcgggca tcctcaaatc catcatgaag aagagagacg gcacacctgg tgcccaaccc 1380
agctccggac ccaagagcct gcagtttgtt ggggtcctca acggagagta cgagagctcc 1440
tccagcgagg acgccagcga cagcgatggc gacagcgaga acggtggcgc cgagcccccg 1500
ggtagctcct cgggctccgg ggatgacagc ggcgggggat ccgactcggg caccccctggc 1560
cctcccagcg gcggggacat ccgggaccct gagcccgagg cggaggcaga gcctcagcag 1620
gtggcacagg ggaggtgcga gctgagcccg cgtctgaggg aggcgtgcgt agcgctgcag 1680
cggcagctga gccggccccg cggagtagcc agcgacggcg gcgcagtgcg cctcgtggcc 1740
caggagtggt ttcgagtgtc cagccagcgg cgctctcagg cggagcccgt ggccaggatg 1800
ctggaagggg tgaggcgcct gggacccgaa ctgctggcgc acgtggtgaa cctggcggat 1860
ggcaacggga acacggccct gcactacagt gtgtcccacg ggaacctggc catcgcaagc 1920
ctgctcctgg atacgggggc ctgcgaggtc aaccgccaga accgagccgg ctactcggcc 1980
ctcatgctgg ctgcactcac ctctgtgagg caggaagagg aggacatggc tgtggtccag 2040
agactcttct gcatgggtga tgtcaatgcc aaggccagtc agacggggca gacagccctc 2100
atgctggcca tcagccatgg ccgacaggac atggtggcaa ccctactggc gtgtggggct 2160
gatgtgaatg cgcaggatgc ggatggggcc acagcgctga tgtgtgccag tgagtatggg 2220
cgcctggaca ccgtgcggct gctgctcacc cagccaggct gtgaccctgc catcctggac 2280
aatgagggca ccagtgccct ggccatcgcc ctggaggctg agcaggatga ggtggccgct 2340
ctgctacatg cccacctgag ctcgggccag cccgacaccc agagcgagtc accccctggc 2400
tcccagacag ccacacctgg tgaaggagaa tgcggtgaca atggagagaa cccccaggtt 2460
cagtaagctg cctcgtctgg ctcactacac ctagctgtgg ggagatctcc tcgtcagtca 2520
cctcagcctt tggcgcacag aagggtccag ggtcccctgc tcagaggcta acactggccg 2580
aagagaaagg caatttcagt tggggtgact gtggcaggaa ggggctcact ctggccccac 2640
caaggtgagg tggggaccaa gtgatagagc cctgatccac ccactctctg aaacttcttt 2700
gctaataaaa 2710

<210> 34
<211> 3527
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3413610CB1

<400> 34
atggccagga gaggtaagaa gcccgtggtg agaacgctgg aggatctgac gctggactcg 60
ggttatggtg gcgcggcgga ctcggtgcgc tcctccaact tgtctttgtg ctgttccgac 120

tcgcacccgg cgtccccgta tggcgggagc tgctggccgc ctctagctga ctccatgcac 180
agccggcaca acagctttga cactgtcaac actgccctgg tggaagactc cgaggggctg 240
gactgcgccg gccagcactg ctcgcggctg ctgccggacc tagacgaggt cccctggact 300
ctccaggagc tggaggcgct gctgctgcgt tcgcgggatc ccgggcagg cccggcggtc 360
cccggcggcc tgcccaagga cgcgctggcc aagctgtcga cgctggtgag ccgggcgctg 420
gtgcgcatag ccaaagaggc gcagcgcctg agcctgcgct tcgccaagtg caccaagtac 480
gagatccaga gcgccatgga gatcgtgctg tcctggggcc tggccgcgca ctgtacggcg 540
gctgcgctgg ccgcactgtc cctctacaac atgagcagcg ccggcggcga ccgcctgggc 600
cgcggcaagt cggcccgctg cggcctcacc ttctccgtgg gccgcgtgta tcgctggatg 660
gtggacagcc gcgtggcgct gcgcatccac gagcacgccg ccatctacct gacagcctgc 720
atggagagcc tcttccggga catctactcg cgggtcgtgg cctccggggt gccccggagc 780
tgcagtggcc ctgggtcagg ctcgggctcc ggcccaggcc cgagctcggg ccctggtgcg 840
gcccccgcgg cggataaaga gcgggaggcg cccgggggag gagcggcgag cggcggcgcc 900
tgcagcgcag ccagcagcgc cagcgggggc agcagctgtt gcgccccgcc ggccgccgcg 960
gccgccgcag tcccgccgac agccgccgcc aaccaccacc atcaccacca ccatgcgctc 1020
cacgaggcgc ccaagttcac cgtggagacc ctggagcaca cggtcaacaa cgactcggag 1080
atctggggtc tcctgcagcc ctaccagcac ctgatctgcg ggaagaacgc cagcggtgac 1140
ctggtgtccc gtgcaatgca tcacctgcag cccctccagg tggaaaggcc cttcctcgtg 1200
ctgccgccgc tgatggagtg gatccgggtg gccgtggcgc acgccggcca ccgccgcagc 1260
ttctccatgg acagcgacga cgtccgccag gcggcccggc tgctgctgcc cggcgtggac 1320
tgcgagccgc gccagctcag ggccgacgac tgcttttgtg catctcgaaa gctggatgcg 1380
gtggccatcg aagccaagtt taagcaggac ctgggtttcc ggatgctgaa ctgtggacga 1440
acagacctgg tgaagcaggc agtgtctctg ctggggcccg atgggatcaa caccatgagc 1500
gaacagggca tgactcccct gatgtatgcc tgcgtccgtg gggacgaggc gatggttcag 1560
atgctgctgg atgccggagc tgacctgaat gtggaggttg tcagtactcc tcataaatat 1620
ccatccgtcc accccgagac ccgccattgg acggctctga cttttgctgt gttgcatgga 1680
catattcctg tagttcagct cctcctggat gctggggcca aggtggaagg ctcagtggag 1740
catggcgagg agaactactc ggaaacaccc ctccagctgg cagctgctgt aggaaatttt 1800
gagctggtta gtttgctgtt ggagcgtggt gccgatcccc tgataggaac catgtacagg 1860
aatggaattt ctacaacccc ccagggtgat atgaactctt tcagccaggc tgcagcccac 1920
ggacacagga atgtgttccg caaactgctc gcccagccag agaaggagaa gagtgatatc 1980
ctgtccctgg aggagattct ggccgagggg actgacctgg cggagacagc cccgcccccc 2040
ttgtgcgcca gccgcaacag caaggccaaa ctgagggccc tgagggaggc catgtatcac 2100
agcgctgagc atggctacgt ggatgtcaca attgatatca ggagcatagg cgtcccgtgg 2160
actctgcaca cgtggctgga gtctttgcgg atcgccttcc agcagcaccg caggcctctc 2220
atccagtgct tgttaaagga gtttaagacc attcaggagg aggaatacac ggaggagctc 2280
gttacccaag gcctgcccct gatgtttgag atcctgaaag cgagcaagaa tgaagtgatc 2340
agccagcagc tgtgcgtcat cttcacacac tgctacgggc cctaccccat ccccaagctc 2400
acagaaatca aacggaaaca gacctcgcgc ttggatcctc atttcttaa caataaagaa 2460
atgtctgatg ttacatttct ggtagaagga agaccatttt atgctcacaa agtgctgtta 2520
tttacagcct ctccaaggtt caaagcactc ctctccagca agccgacaaa tgatggcacc 2580
tgcatagaga ttggttatgt gaaatactcc atctttcagc tggttatgca gtatctctac 2640
tatggtggcc cagagtcact gctcattaaa aacaatgaga tcatggagct tctgtctgct 2700
gctaagtttt tccagctgga ggctttgcag cgacactgtg agattatctg tgcgaaaagc 2760
atcaataccg acaactgtgt ggatatttac aaccatgcca agtttcttgg agtcacagag 2820
ctctcagcat attgcgaagg ctactttctc aaaaacatga tggtcctcat tgaaaacgaa 2880
gcattcaagc agctcctgta tgacaaaaat ggtgaaggga ccggccagga tgtgctccag 2940
gacttacaga ggacgttggc catcagaatt cagtccatcc acttgtcgtc ttccaaaggt 3000
tccgtggtat gaaacgccta gtgcagggaa tgcttcccgg gaactttcca gttctcctgc 3060
cgcattggct ttacacaaac acagacaaat tccacctggc acctgttttt ggctgggcca 3120
aggagctgcc tctactgctc ccacgtgttc ctgttgaaaa acaaaggact ttccactggt 3180
ctgcagatca gatcagctgg gtccagagtt taatgggcaa ctggacaacc aagttaaccc 3240
caattgaaag caccccctagg accattgaac acccactgcc ggggaccact gtccagtgaa 3300
tggattgagg ccttttaaag gtcactcagg ttccaggttg acagttggag gacttcaccg 3360
taccaaccct gaagagattg tattacacat taaggacctt ggtagctgtg cttcagcaaa 3420
cgtcaaccat ggtagcaaat tggtgaggct gtgaccaata atgaggaaat aatctggcaa 3480

atttttaggg gtgggaactt ttttaaatgt tcatttaaaa aaaaaaa 3527

<210> 35
<211> 3251
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3276394CB1

<400> 35
cggcgtcaga gacactgcga gcggcgagcg cggtggggcc gcatctgcat cagccgccgc 60
agccgctgcg gggccgcgaa caaagaggag gagccgaggc gcgagagcaa agtctgaaat 120
ggatgttaca tgagtcattt taagggatgc acacaactat gaacatttct gaagattttt 180
tctcagtaaa gtagataaag atggatgaat cagccttgtt ggatcttttg gagtgtccgg 240
tgtgtctaga gcgccttgat gcttctgcga aggtcttgcc ttgccagcat acgttttgca 300
agcgatgttt gctggggatc gtaggttctc gaaatgaact cagatgtccc gagtgcagga 360
ctcttgttgg ctcgggtgtc gaggagcttc ccagtaacat cttgctggtc agacttctgg 420
atggcatcaa acagaggcct tggaaacctg gtcctggtgg gggaagtggg accaactgca 480
caaatgcatt aaggtctcag agcagcactg tggctaattg tagctcaaaa gatctgcaga 540
gctcccaggg cggacagcag cctcgggtgc aatcctggag ccccccagtg aggggtatac 600
ctcagttacc atgtgccaaa gcattataca actatgaagg aaaagagcct ggagacctta 660
aattcagcaa aggtgacatc atcattttgc gaagacaagt ggatgaaaat tggtaccatg 720
gggaagtcaa tggaatccat ggcttttttc ccaccaactt tgtgcagatt attaaaccgt 780
tacctcagcc cccatctcag tgcaaagcac tttatgactt tgaagtgaaa gacaaggaag 840
cagacaaaga ttgccttcca tttgcaaagg atgatgttct gactgtgatc cgaagagtgg 900
atgaaaactg ggctgaagga atgctggcag acaaaatagg aatatttcca atttcatatg 960
ttgagtttaa ctcggctgct aagcagctga tagaatggga taagcctcct gtgccaggag 1020
ttgatgctgg agaatgttcc tcggcagcag cccagagcag cactgcccca aagcactccg 1080
acaccaagaa gaacaccaaa aagcggcact ccttcacttc cctcactatg gccaacaagt 1140
cctcccaggc atcccagaac cgccactcca tggagatcag ccccccctgtc ctcatcagct 1200
ccagcaaccc cactgctgct gcacggatca gcgagctgtc tgggctctcc tgcagtgccc 1260
cttctcaggt tcatataagt accaccgggt taattgtgac cccgccccca agcagcccag 1320
tgacaactgg cccctcgttt actttcccat cagatgttcc ctaccaagct gcccttggaa 1380
ctttgaatcc tcctcttcca ccacccccctc tcctggctgc cactgtcctt gcctccacac 1440
caccaggcgc caccgccgct gctgctgctg ctggaatggg accgaggccc atggcaggat 1500
ccactgacca gattgcacat ttacggccgc agactcgccc cagtgtgtat gttgctatat 1560
atccatacac tcctcggaaa gaggatgaac tagagctgag aaaaggggag atgtttttag 1620
tgtttgagcg ctgccaggat ggctggttca aagggacatc catgcatacc agcaagatag 1680
ggggttttccc tggcaattat gtggcaccag tcacaagggc ggtgacaaat gcttcccaag 1740
ctaaagtccc tatgtctaca gctggccaga caagtcgggg agtgaccatg gtcagtcctt 1800
ccacggcagg agggcctgcc cagaagctcc agggaaatgg cgtggctggg agtcccagtg 1860
ttgtccccgc agctgtggta tcagcagctc acatccagac aagtcctcag gctaaggtct 1920
tgttgcacat gacggggcaa atgacagtca accaggcccg caatgctgtg aggacagttg 1980
cagcgcacaa ccaggaacgc ccccacggcag cagtgacacc catccaggta cagaatgccg 2040
ccggcctcag ccctgcatct gtgggcctgt cccatcactc gctggcctcc ccacaacctg 2100
cgcctctgat gccaggctca gccacgcaca ctgctgccat cagtatcagt cgagccagtg 2160
cccctctggc ctgtgcagca gctgctccac tgacttcccc aagcatcacc agtgcttctc 2220
tggaggctga gcccagtggc cggatagtga ccgttctccc tggactcccc acatctcctg 2280
acagtgcttc atcagcttgt gggaacagtt cagcaaccaa accagacaag gatagcaaaa 2340
aagaaaaaaa gggtttgttg aagttgcttt ctggcgcctc cactaaacgg aagccccgcg 2400
tgtctcctcc agcatcgccc accctagaag tggagctggg cagtgcagag cttcctctcc 2460
agggagcggt ggggcccgaa ctgccaccag gaggtggcca tggcagggca ggctcctgcc 2520
ctgtggacgg ggacggaccg gtcacgactg cagtggcagg agcagccctg gcccaggatg 2580
cttttcatag gaaggcaagt tccctggact ccgcagttcc catcgctcca cctcctcgcc 2640

aggcctgttc ctccctgggt cctgtcttga atgagtctag acctgtcgtt tgtgaaaggc 2700
acagggtggt ggtttcctat cctcctcaga gtgaggcaga acttgaactt aaagaaggag 2760
atattgtgtt tgttcataaa aaacgagagg atggctggtt caaaggcaca ttacaacgta 2820
atgggaaaac tggccttttc ccaggaagct ttgtggaaaa catatgagga gactgacact 2880
gaagaagctt aaaatcactt cacacaacaa agtagcacaa agcagtttaa cagaaagagc 2940
acatttgtgg acttccagat ggtcaggaga tgagcaaagg attggtatgt gactctgatg 3000
ccccagcaca gttaccccag cgagcagagt gaagaagatg tttgtgtggg ttttgttagt 3060
ctggattcgg atgtataagg tgtgccttgt actgtctgat ttactacaca gagaaacttt 3120
ttttttttt aaagattttt gactaaagtg gccgattgtt ttccgggtta actaatttat 3180
tgggttttta acttgaactt tcggtaaaaa aaaaagctgg ggaaatggtt tggaaatttt 3240
attttgaaag g 3251

<210> 36
<211> 1600
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7602049CB1

<400> 36
gctccacgca gcccggctgg gcagcaaggg acagaacaga ggcggccgct gacagcacca 60
gcatgtctta cagtgtgacc ctgactgggc ccgggccctg gggcttccgt ctgcaggggg 120
gcaaggactt caacatgccc ctcactatct cccggatcac accaggcagc aaggcagccc 180
agtcccagct cagccagggt gacctcgtgg tggccattga cggcgtcaac acagacacca 240
tgacccacct ggaagcccag aacaagatca agtctgccag ctacaacttg agcctcaccc 300
tgcagaaatc aaagcgtccc attcccatct ccacgacagc acctccagtc cagacccctc 360
tgccggtgat ccctcaccag aaggtggtag tcaactctcc agccaacgcc gactaccagg 420
aacgcttcaa ccccagtgcc ctgaaggact cggccctgtc cacccacaag cccatcgagg 480
tgaaggggct gggcggcaag gccaccatca tccatgcgca gtacaacacg cccatcagca 540
tgtattccca ggatgccatc atggatgcca tcgctgggca ggcccaagcc caaggcagtg 600
acttcagtgg gagcctccct attaaggacc ttgccgtaga cagcgcctct cccgtctacc 660
aggctgtgat taagagccag aacaagccag aagatgaggc tgacgagtgg gcacgccgtt 720
cctccaacct gcagtctcgc tccttccgca tcctggccca gatgacgggg acagaattca 780
tgcaagaccc tgatgaagaa gctctgcgaa ggtcaaggga aaggtttgaa acggaacgta 840
acagcccacg ttttgccaaa ttgcgcaact ggcaccatgg cctttcagcc caaatcctta 900
atgttaaaag ctaaaaggct gcctggaatc cccccacccc aacaggctgg actccctcca 960
tccttacccc cacacagatc tggcatgtga gccccacggt gatgcttgac aatgtataac 1020
tctgctgggg gcacctctga tggccaaccg cagcatttct gtcctctgcc caccccagag 1080
ctgatgctgg ggcccagccc cctgcagctc tgtacccacc aaacctcccc agggcaaccc 1140
tcgccacccc ccaaatagcc cgtagcccaa tcccctgccc tctgcacagg gccttagctg 1200
tagaccagag agggcaggag gggtttgctg gcataacacc ccagaaccaa gggaaatgga 1260
tgggccgctg ctcagtttcc caccatcctc agctcctggc ctcatcccct cctagaatga 1320
gtcacccgta gatcagggtc tggggaagag gctgatccct ggcgctgccc ggctccctcg 1380
ctgccctctg gagctcaggg cagcccggaa tagggctctt tgaagaggaa gtagaagccc 1440
cagggtaatg aggcagagac ccctcctggc agtggtgagg tgggggcatg caccctcctt 1500
tctgtaccgt gtgtgctggc tccatagttc tctcttctgt acatataagc atgcttgttc 1560
tgaaataaag aagatttgaa gtgaaccaca aaaaaaaaa 1600